

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:17:24 ; Search time 183 Seconds
(without alignments)
48.020 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116
Sequence: 1 QEANYAGNQSIGNYRGWNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	20	7 ADD93642	Add93642 Streptoco
2	116	100.0	20	9 ADX37265	Adx37265 Streptoco
3	116	100.0	431	7 ADD93649	Add93649 Streptoco
4	116	100.0	431	7 ADD93650	Add93650 Streptoco
5	116	100.0	431	7 ADD93653	Add93653 Streptoco
6	116	100.0	431	9 ADX37272	Adx37272 Streptoco
7	116	100.0	431	9 ADX37273	Adx37273 Streptoco
8	116	100.0	431	9 ADX37276	Adx37276 Streptoco
9	116	100.0	431	9 AEB91500	Aeb91500 Microbial
10	116	100.0	432	7 ADD93651	Add93651 Streptoco
11	116	100.0	432	7 ADD93652	Add93652 Streptoco
12	116	100.0	432	9 ADX37274	Adx37274 Streptoco
13	116	100.0	432	9 ADX37275	Adx37275 Streptoco
14	96	82.8	132	2 AAW60944	Aaw60944 Streptoco
15	96	82.8	392	6 ABU02747	Abu02747 S. pneumo
16	96	82.8	392	8 ADK47859	Adk47859 Streptoco
17	96	82.8	392	8 ADT50227	Adt50227 S.pneumon
18	96	82.8	392	8 ADT50226	Adt50226 S.pneumon
19	96	82.8	392	8 ADT50165	Adt50165 S.pneumon
20	96	82.8	399	8 ADT50165	Adt50165 S.pneumon
21	96	82.8	399	9 AEA58465	Aea58465 Streptoco
22	89	76.7	395	5 ABP25918	Abp25918 Streptoco
23	89	76.7	447	5 ABP29684	Abp29684 Streptoco
24	89	76.7	447	8 ADU69524	Adu69524 S agalact

25	89	76.7	447	8	ADV88392	Adv88392 Streptoco
26	89	76.7	447	8	ADV81808	Adv81808 Streptoco
27	89	76.7	447	8	ADV79645	Adv79645 Streptoco
28	84	72.4	165	5	ABP29842	Abp29842 Streptoco
29	84	72.4	169	5	ABP27864	Abp27864 Streptoco
30	84	72.4	169	8	ADU69652	Adu69652 S agalact
31	84	72.4	169	8	ADU69652	Adu69652 S agalact
32	84	72.4	169	8	ADV88318	Adv88318 Streptoco
33	84	72.4	169	8	ADV81739	Adv81739 Streptoco
34	84	72.4	169	8	ADV79571	Adv79571 Streptoco
35	77	66.4	398	5	ABP25919	Abp25919 Streptoco
36	77	66.4	398	8	ADR83884	Adr83884 S. pyogen
37	77	66.4	398	8	AEB91542	Aeb91542 Microbial
38	70	60.3	211	9	AEB91642	Aeb91642 Microbial
39	50	43.1	133	4	AAU19320	Aau19320 Human G p
40	50	43.1	510	9	ARM94545	Arm94545 M. xanthu
41	49	42.2	123	3	AAV87588	Aav87588 Aniopept
42	49	42.2	308	3	AAV87586	Aav87586 Aeromonas
43	49	42.2	407	3	AAV87585	Aav87585 Aeromonas
44	49	42.2	491	3	AAV87584	Aav87584 Aeromonas
45	49	42.2	590	3	AAV87583	Aav87583 Aeromonas

ALIGNMENTS

RESULT 1
ADD93642
ID ADD93642 standard; peptide; 20 AA.
XX
AC ADD93642;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B peptide fragment.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US0006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
XX (FORS-) FORSYTH INST.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX
XX Claim 11; Page 10; 49pp; English.

The present sequence is that of a peptide comprising amino acid residues 403-422 of the glucan binding protein-B (GbpB) of Streptococcus mutans strain SK32 ADD93649. The peptide binds to a major histocompatibility complex (MHC) class II protein. It was identified as a potential B cell epitope using a matrix-based algorithm for epitope prediction, which was used to search the primary amino acid sequence of GbpB for known MHC class II binding motifs. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These comprise MHC class II protein-binding GbpB peptides covalently linked with peptide subunits (preferably from the catalytic domain) of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiptopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

SQ Sequence 20 AA;
 Query Match 100.0%; Score 116; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QEANYAGNQSIGNRGWFNP 20
 |||||
 Db 1 QEANYAGNQSIGNRGWFNP 20
 |||||

RESULT 2
 ADX37265
 ID ADX37265 standard; peptide; 20 AA.
 XX
 AC ADX37265;
 XX
 DT 21-APR-2005 (first entry)
 XX
 DE Streptococcus mutant glucan binding protein B peptide #22.
 XX
 KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX
 OS Streptococcus mutans.
 XX
 PN US2005031633-A1.
 XX
 PD 10-FEB-2005.

PF 09-MAR-2004; 2004US-00797821.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 98US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.
 XX
 PA (SMIT/) SMITH D J.
 PA (TAUB/) TAUBMAN M A.
 XX
 PI Smith DJ, Taubman MA;
 XX
 DR WPI; 2005-151644/16.
 XX
 PT New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.

PS Claim 4; SEQ ID NO 22; 73pp; English.
 XX
 CC The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB-derived peptide of the invention.

SQ Sequence 20 AA;
 Query Match 100.0%; Score 116; DB 9; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QEANYAGNQSIGNRGWFNP 20
 |||||
 Db 1 QEANYAGNQSIGNRGWFNP 20
 |||||

RESULT 3
 ADD93649
 ID ADD93649 standard; protein; 431 AA.
 XX
 AC ADD93649;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Streptococcus mutans glucan binding protein-B.
 XX
 KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
 XX
 OS Streptococcus mutans.
 XX
 FH Key Location/Qualifiers
 FT Region 6..25
 FT /note= "HLA-binding peptide"
 FT Region 16..35
 FT /note= "HLA-binding peptide"
 FT Region 33..52
 FT /note= "HLA-binding peptide"
 FT Region 37..56
 FT /note= "HLA-binding peptide"
 FT Region 48..67
 FT /note= "HLA-binding peptide"
 FT Region 52..71
 FT /note= "HLA-binding peptide"
 FT Region 88..107
 FT /note= "HLA-binding peptide"
 FT Region 113..132
 FT /note= "HLA-binding peptide"
 FT Region 117..136
 FT /note= "HLA-binding peptide"
 FT Region 137..156
 FT /note= "HLA-binding peptide"
 FT Region 174..193
 FT /note= "HLA-binding peptide"
 FT Region 194..213
 FT /note= "HLA-binding peptide"
 FT Region 214..233
 FT /note= "HLA-binding peptide"
 FT Region 248..267
 FT /note= "HLA-binding peptide"
 FT Region 289..308
 FT /note= "HLA-binding peptide"
 FT Region 306..325
 FT /note= "HLA-binding peptide"
 FT Region 311..330
 FT /note= "HLA-binding peptide"
 FT Region 349..368
 FT /note= "HLA-binding peptide"
 FT Region 365..384
 FT /note= "HLA-binding peptide"
 FT Region 383..402
 FT /note= "HLA-binding peptide"
 FT Region 403..422
 FT /note= "HLA-binding peptide"
 XX
 WO2003075845-A2.
 PD 18-SEP-2003.
 XX
 PF 07-MAR-2003; 2003WO-US006962.
 XX
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 XX
 PA (FORS-) FORSYTH INST.
 XX
 PI Smith DJ, Taubman MA;
 XX

```

DR WPI; 2003-845091/78.
DR GENBANK; AY046410.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 6; Page 7; 49pp; English.
XX
CC The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 116; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNVRGWFNP 20
DB 403 QEANYAGNQSIGNVRGWFNP 422

RESULT 4
ADD93650
ID ADD93650 standard; protein; 431 AA.
XX
AC ADD93650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2003-845091/78.
DR GENBANK; AY046411.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 5; Page 8; 49pp; English.
XX
CC The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 3VP4. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental

```

```

CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 116; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNVRGWFNP 20
DB 403 QEANYAGNQSIGNVRGWFNP 422

RESULT 5
ADD93653
ID ADD93653 standard; protein; 431 AA.
XX
AC ADD93653;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2003-845091/78.
DR GENBANK; AY046414.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 5; Page 8-9; 49pp; English.
XX
CC The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 5M3. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 116; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;

```



```

PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 3; SEQ ID NO 33; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 431 AA;
XX
Query Match 100.0%; Score 116; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNRGWFNP 20
Db |||||
403 QEANYAGNQSIGNRGWFNP 422

RESULT 9
AEB91500
ID AEB91500 standard; protein; 431 AA.
XX
AC AEB91500;
XX
DT 20-OCT-2005 (first entry)
XX
DE Microbial pathogen adhesin protein sequence, SEQ ID NO:210.
XX
KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;
KW bordetella pertussis infection; antibacterial; pneumonia;
KW antiinflammatory; respiratory-gen.; gastric ulcer; antitumor;
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
XX
XX Streptococcus mutans.
OS
XX
XX WO2005076010-A2.
PN
XX
XX 18-AUG-2005.
PD
XX
XX 07-FEB-2005; 2005WO-IN000037.
PF
XX
XX 06-FEB-2004; 2004IN-DE000173.
PR
XX 20-JUL-2004; 2004US-0589227P.
PR
XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.
PA
XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;
PI
XX WPI; 2005-597835/61.
XX
XX

```

```

PT Computational method for identifying adhesin and adhesin like molecules,
PT comprises computing sequence-based attributes of protein sequences using
XX neural network software and training an artificial neural network.
XX
PS Claim 16; SEQ ID NO 210; 402pp; English.
XX
XX The present invention relates to a computational method (M1) for
CC identifying adhesin and adhesin-like proteins, by computing the sequence-
CC based attributes of protein sequences using five attribute modules of a
CC neural network software, training an artificial neural network (ANN) for
CC each of the computed five attributes, and identifying the adhesin and
CC adhesin-like proteins having probability of being an adhesin (Pad) as
CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
CC genes encoding adhesin and adhesin-like proteins, having 105 fully
CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
CC 280-384) sequences; and a fully connected multilayer feed forward ANN (1)
CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
CC proteins, of therapeutic potential, and identifying and short-listing
CC proteins, for further testing in development of new vaccine formulations
CC to eliminate diseases caused by various pathogenic organisms. (M1) is
CC useful for identifying putative adhesins that are important in drug
CC discovery and preventing therapeutics for whooping cough, pneumonia,
CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from
CC distantly related organisms, and from bacteria belonging to a wide
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
CC unique proteins. The present sequence is a microbial pathogen adhesin
CC protein sequence.
XX
XX Sequence 431 AA;
XX
Query Match 100.0%; Score 116; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNRGWFNP 20
Db |||||
403 QEANYAGNQSIGNRGWFNP 422

RESULT 10
ADD93651
ID ADD93651 standard; protein; 432 AA.
XX
AC ADD93651;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
XX Streptococcus mutans.
OS
XX
XX WO2003075845-A2.
PN
XX
XX 18-SEP-2003.
PD
XX
XX 07-MAR-2003; 2003WO-US006962.
PF
XX
XX 07-MAR-2002; 2002US-0363209P.
PR
XX 08-AUG-2002; 2002US-0402483P.
PR
XX (FORS-) FORSYTH INST.
PA
XX Smith DJ, Taubman MA;
PI
XX WPI; 2003-845091/78.
XX
XX GENBANK; AY046412.
DR

```


QY 1 QEANYAGNOSIGNYRGWFNP 20
 Db 404 QEANYAGNOSIGNYRGWFNP 423

RESULT 13

ADX37275
 ID ADX37275 standard; protein; 432 AA.

XX AC ADX37275;
 XX 21-APR-2005 (first entry)
 XX Streptococcus mutant glucan binding protein B variant #4.
 DE immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 OS Streptococcus mutans.

XX US2005031633-A1.
 XX 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.
 XX 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 99US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.
 PA (TAUB/) TAUBMAN M A.
 PA Smith DJ, Taubman MA;
 PI WPI; 2005-151644/16.
 DR New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.

XX Claim 3; SEQ ID NO 32; 73pp; English.
 XX The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 432 AA;

Query Match 100.0%; Score 116; DB 9; Length 432;
 Best Local Similarity 100.0%; Pred. No. 8.6e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNOSIGNYRGWFNP 20
 Db 404 QEANYAGNOSIGNYRGWFNP 423

RESULT 14

AAW60944
 ID AAW60944 standard; protein; 132 AA.

XX AAW60944;
 XX 13-OCT-1998 (first entry)

XX Streptococcus pneumoniae; type 4 strain.

PN WO200277021-A2.

DE Streptococcus pneumoniae encoded polypeptide.

XX coding region; ORF; open reading frame; antibacterial; infection;
 KW prevention; meningitis.

XX Streptococcus pneumoniae.

XX WO9819689-A1.

XX 14-MAY-1998.

XX 27-OCT-1997; 97WO-US019226.

XX 01-NOV-1996; 96US-0029930P.

XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Reid RH, Zarfos PN;
 XX WPI; 1998-286586/25.
 DR N-PSDB; AAV37344.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful, e.g.
 PT for identifying anti-bacterial(s) for treatment and prevention of
 PT meningitis.

XX Claim 11; Page 47; 130pp; English.

XX The sequence is that of the polypeptide encoded by a region isolated from
 CC S. pneumoniae. The protein, or agonists of it, may be useful as an
 CC antibacterial for treatment or prevention of infection, specifically
 CC caused by S. pneumoniae (particularly meningitis) but possibly also
 CC Helicobacter pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other invasive
 CC procedure. The protein, or nucleic acid encoding it, can also be used in
 CC vaccines to induce a cellular and/or humoral immune response, or to
 CC screen for other antibacterials. The DNA may also contain flanking
 CC sequences that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein can be used
 CC diagnostically, e.g. to detect a mutation for serotyping or classifying
 CC infectious agents

XX Sequence 132 AA;

Query Match 82.8%; Score 96; DB 2; Length 132;
 Best Local Similarity 78.9%; Pred. No. 3.5e-07;
 Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
 Db 100 ESNYAGNRTIGNHRCWFNP 118

RESULT 15

ABU02747
 ID ABU02747 standard; protein; 392 AA.

XX AC ABU02747;

XX 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain protein from coding region #2326.

XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.

XX Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

XX 03-OCT-2002.
 PD
 XX 27-MAR-2002; 2002WO-IB002163.
 PF
 XX 27-MAR-2001; 2001GB-00007658.
 PR
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI: 2003-040579/03.
 DR N-PSDB; ABX08038.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 1; SEQ ID NO 4652; 56pp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 392 AA;

Query Match 82.8%; Score 96; DB 6; Length 392;
 Best Local Similarity 78.9%; Pred. No. 1.3e-06;
 Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 2 EANYAGNOSIGNRGWFNP 20
 |:|||||:|||||
 Db 360 ESNYAGNRTIGNRGWFNP 378

Search completed: February 15, 2006, 22:20:44
 Job time : 185 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:20:59 ; Search time 38 Seconds
(without alignments)
50.640 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNQSIGNYRGWFNP 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	82.8	392	2 G95258	secreted 45 kd pro
2	96	82.8	392	2 B98124	general stress pro
3	55	47.4	332	2 AC0260	hypothetical phage
4	50	43.1	1061	2 A40609	OmpA-related prote
5	47	40.5	261	2 S71364	ubiquinol-cytochro
6	47	40.5	837	1 S54429	outer membrane ush
7	46	39.7	319	1 S65523	dihydrodipicolinat
8	46	39.7	510	2 A98154	hypothetical prote
9	46	39.7	521	2 AC3134	hypothetical prote
10	46	39.7	619	1 KSNCL0	laccase (EC 1.10.3
11	46	39.7	619	1 KSNCLT	laccase (EC 1.10.3
12	46	39.7	888	2 C90595	hypothetical prote
13	45.5	39.2	350	2 AG3425	iolE protein limpo
14	45.5	39.2	350	2 AE2954	conserved hypotet
15	45.5	39.2	370	2 H98328	hypothetical prote
16	45.5	39.2	385	2 D41732	heterogeneous nucl
17	45	38.8	108	1 XVM573	Ig kappa chain v r
18	45	38.8	242	1 F75433	probable phosphoes
19	45	38.8	514	2 B64634	hypothetical prote
20	45	38.8	514	2 E71880	probable outer mem
21	45	38.8	536	2 T35109	hypothetical prote
22	45	38.8	621	2 S72493	laccase (EC 1.10.3
23	45	38.8	714	2 AT6590	probable chemotaxi
24	45	38.8	835	2 T06590	probable beta-gala
25	44.5	38.4	1059	2 E87058	isoleucyl-tRNA syn
26	44	37.9	43	2 S21065	Ig kappa chain v r
27	44	37.9	147	2 I40625	hypothetical prote
28	44	37.9	237	2 F97109	uncharacterized pr
29	44	37.9	416	2 S09334	transcobalamin I p

30	44	37.9	645	2 T16078	hypothetical prote
31	44	37.9	1006	2 S20126	exoribonuclease RA
32	43.5	37.5	160	2 JC7622	actinohivin precu
33	43.5	37.5	430	2 S55325	endo-beta-1,6-gluc
34	43	37.1	125	2 S04936	Ig kappa chain pre
35	43	37.1	130	2 PL0113	Ig kappa chain pre
36	43	37.1	137	2 S50009	hypothetical prote
37	43	37.1	177	2 H81743	conserved hypotet
38	43	37.1	340	2 T26526	hypothetical prote
39	43	37.1	372	2 AG2827	HFLK protein limpo
40	43	37.1	373	2 E97605	proteinase chain h
41	43	37.1	451	1 S27617	glucarate dehydrat
42	43	37.1	516	2 T15633	hypothetical prote
43	43	37.1	548	2 T27542	hypothetical prote
44	43	37.1	614	2 S78277	dnak-type molecula
45	43	37.1	626	2 E84432	hypothetical prote

ALIGNMENTS

RESULT 1

G95258

secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004

C:Accession: G95258

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
non, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95258

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AB005672; I

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP2216

Query Match 82.8%; Score 96; DB 2; Length 392;

Best Local Similarity 78.9%; Pred. No. 6.3e-07;

Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNQSIGNYRGWFNP 20

I:|||||:|||||

DB 360 ESNYAGNRTIGNHRGWFP 378

RESULT 2

B98124

General stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004

C:Accession: B98124

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Es
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98124

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AB007317; I

C:Genetics:

A:Gene: gsp-781

Query Match 82.8%; Score 96; DB 2; Length 392;
 Best Local Similarity 78.9%; Pred. No. 6.3e-07;
 Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFP 20
 :|||||:|||||
 Db 360 ESNYAGNTGNHRCWFP 378
 :|||||:|||||

RESULT 3
 AC0260
 hypochetrical phage protein YPO2132 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AC0260
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AC0260
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-332 <UNP>
 A:CROSS-references: UNIPROT:Q8ZEM9; UNIPARC:UPI00000CD8A6; GB:AL590842; PIDN:CAC90943.1;
 C:Genetics:
 A:Gene: YPO2132

Query Match 47.4%; Score 55; DB 2; Length 332;
 Best Local Similarity 69.2%; Pred. No. 0.94;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGNQSIGNRGWF 18
 :|||:|||||
 Db 275 AGNINNGYRGWY 287
 :|||:|||||

RESULT 4
 A40609
 OmpA-related protein Oar - Myxococcus xanthus
 C:Species: Myxococcus xanthus
 C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C:Accession: A40609
 R:Martinez-Canamero, M.; Munoz-Dorado, J.; Farez-Vidal, E.; Inouye, M.; Inouye, S.
 J. Bacteriol. 175, 4756-4763, 1993
 A:Title: Oar, a 115-kilodalton membrane protein required for development of Myxococcus
 A:Reference number: A40609; MUID:93328680; PMID:8335633
 A:Accession: A40609
 A:Status: preliminary
 A:Molecule type: DNA; protein
 A:Residues: 1-1061 <MAR>
 A:CROSS-references: UNIPROT:P38370; UNIPARC:UPI0000130B68; GB:S64103; NID:g402803; PIDN:
 A:Note: sequence extracted from NCBI backbone (NCBIN:135644, NCBI:P:135645)

Query Match 43.1%; Score 50; DB 2; Length 1061;
 Best Local Similarity 47.4%; Pred. No. 18;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFP 20
 :|||:|||||
 Db 844 QANYTWSRLYGNYPGLFRP 862
 :|||:|||||

RESULT 5
 S71364
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) Rieske iron-sulfur protein precursor - Ch
 N:Alternate names: Rieske iron-sulfur protein
 C:Species: Chlamydomonas reinhardtii
 C>Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 09-Jul-2004
 C:Accession: S71364; S78209
 R:Atteia, A.; Franzen, L.G.
 Eur. J. Biochem. 237, 792-799, 1996

A:Title: Identification, cDNA sequence and deduced amino acid sequence of the mitochondri
 g and subunit interaction.
 A:Reference number: S71364; MUID:96235247; PMID:8647127
 A:Accession: S71364
 A:Molecule type: mRNA
 A:Residues: 1-261 <ATT>
 A:CROSS-references: UNIPROT:Q42703; UNIPARC:UPI000009F24B; EMBL:X91795; NID:g1418673; PII
 A:Experimental source: strain CW15
 A:Accession: S78209
 A:Molecule type: protein
 A:Residues: 55-74 <ATY>
 A:CROSS-references: UNIPARC:UPI0000174C82
 C:Genetics:
 A:Genome: nuclear
 C:Superfamily: ubiquinol-cytochrome-c reductase (cytochrome c1) (complex III), iron-sulf
 C:Keywords: 2Fe-2S; electron transfer; membrane-associated complex; metalloprotein; mito
 F1-54/Domain: transit peptide (mitochondrion) #status predicted <TN>
 F:55-261/Product: ubiquinol-cytochrome-c reductase iron-sulfur protein #status experiment
 F:194-241/Domain: Rieske [2Fe-2S] homology <RSK>
 F:204-206, 223, 226/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #status p
 F:209-225/Disulfide bonds: #status predicted
 F:226/Active site: His #status predicted

Query Match 40.5%; Score 47; DB 2; Length 261;
 Best Local Similarity 77.8%; Pred. No. 12;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 12 GNYRGWFP 20
 :|||:|||||
 Db 216 GNYQGWFCP 224
 :|||:|||||

RESULT 6
 S54429
 outer membrane usher protein hifc precursor - Haemophilus influenzae (strain AM30)
 C:Species: Haemophilus influenzae
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C:Accession: S54429
 R:van Ham, S.M.; van Alphen, L.; Mooi, F.R.; van Putten, J.P.M.
 Mol. Microbiol. 13, 673-684, 1994
 A:Title: The fimbrial gene cluster of Haemophilus influenzae type b.
 A:Reference number: S54428; MUID:95089703; PMID:7997179
 A:Accession: S54429
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-837 <VAN>
 A:CROSS-references: UNIPROT:P45997; UNIPARC:UPI000012C67D; EMBL:Z33502; NID:G535165; PIDN
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
 C:Genetics:
 A:Gene: hifc
 A:Start codon: GTG
 C:Superfamily: outer membrane usher protein fimD
 C:Keywords: membrane protein

Query Match 40.5%; Score 47; DB 1; Length 837;
 Best Local Similarity 50.0%; Pred. No. 41;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFP 19
 :|||:|||||
 Db 623 DANLAHNSIGSYRASYS 640
 :|||:|||||

RESULT 7
 S56523
 dihydrodipicolinate synthase homolog yjH - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
 C:Accession: S56523; D65243
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.1
 A:Reference number: S56314; MUID:95334362; PMID:7610040

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AC3134

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-521 <KUR>

A;Cross-references: UNIPROT:Q8UGW1; UNIPARC:UPI00000D267E; GB:AE008689; PIDN:AAL45489.1;

C;Genetics:

A;Gene: Atu4695

A;Map position: linear chromosome

Query Match 39.7%; Score 46; DB 2; Length 521;

Best Local Similarity 50.0%; Pred. No. 35;

Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GNQSIGNYRGWFNP 20

Db 437 GAMGVSNYGVWSNP 450

RESULT 10

KSNCLO

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)

N;Alternate names: urishiol oxidase

C;Species: Neurospora crassa

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004

C;Accession: A28523; A29762

R;Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.

J. Biol. Chem. 263, 885-896, 1988

A;Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and

A;Reference number: A28523; MUID:88087214; PMID:2961749

A;Accession: A28523

A;Molecule type: DNA

A;Residues: 1-619 <GER>

A;Cross-references: UNIPROT:P06811; UNIPARC:UPI00001721B1; EMBL:M14554

R;Germann, U.A.; Lerch, K.

Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986

A;Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospora

A;Reference number: A29762; MUID:87067412; PMID:2947240

A;Accession: A29762

A;Molecule type: DNA

A;Residues: 379-619 <GE2>

A;Cross-references: UNIPARC:UPI0000168938; GB:M14554; NID:G168823; PIDN:AAA33590.1; PID:9

C;Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone

C;Genetics:

A;Introns: 86/3

C;Superfamily: laccase

C;Keywords: copper; glycoprotein; oxidoreductase

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-49/Domain: propeptide #status predicted <PRO>

F;50-619/Product: laccase #status predicted <MAT>

F;84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>

F;216-372/Domain: middle beta-barrel #status predicted <BB2>

F;431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>

F;139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;144,480/Binding site: copper (His) (type 2) #status predicted

F;146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted

F;477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 39.7%; Score 46; DB 1; Length 619;

Best Local Similarity 66.7%; Pred. No. 42;

Matches 8; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

Qy 8 NQSIGNYRGWFNP 19

Db 331 NQAVGVN--WFN 340

RESULT 11

KSNCLO

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)

N;Alternate names: urishiol oxidase

C;Species: Neurospora crassa

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
 C/Accession: B28523
 J/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A/Title: Characterization of two allelic forms of *Neurospora crassa* laccase. Amino- and
 A/Reference number: A28523; MUID:88087314; PMID:2961749
 A/Accession: B28523
 A/Molecule type: DNA
 A/Residues: 1-619 <GER>
 A/Cross-references: UNIPROT:P10574; UNIPARC:UPI000012E1B2; EMBL:M18334; NID:G168827; PID
 C/Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
 C/Genetics:
 A/Introns: 86/3
 C/Superfamily: laccase
 C/Keywords: copper; glycoprotein; oxidoreductase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-49/Domain: propeptide #status predicted <PRO>
 F:50-619/Product: laccase #status predicted <MAT>
 F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
 F:216-372/Domain: middle beta-barrel #status predicted <BB2>
 F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
 F:139,282,295,340,422,444/Binding site: carbohydrate (Aen) (covalent) #status predicted
 F:144,480/Binding site: copper (His) (type 2) #status predicted
 F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred
 F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 39.7%; Score 46; DB 1; Length 619;
 Best Local Similarity 66.7%; Pred. No. 42;
 Matches 8; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 8 NOSIGNYRGWEN 19
 ||:|||||
 DB 331 NOAVGNV--WFN 340

RESULT 12
 C90595
 hypothetical protein MYPV_6670 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C/Species: Mycoplasma pulmonis
 C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 R/Accession: C90595
 R/Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
 Nucleic Acids Res. 29, 2145-2153, 2001
 A/Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulm*
 A/Reference number: A99512; MUID:21267165; PMID:11353084
 A/Accession: C90595
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-888 <KUR>
 A/Cross-references: UNIPROT:Q98PQ2; UNIPARC:UPI00000D463E; GB:AL445566; PID:G14090082; F
 A/Experimental source: strain UAB CTIP
 C/Genetics:
 A/Gene: MYPV_6670
 A/Genetic code: SGC3
 C/Superfamily: isoleucine-tRNA ligase

Query Match 39.7%; Score 46; DB 2; Length 888;
 Best Local Similarity 63.6%; Pred. No. 61;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 QSIGNYRGWEN 19
 :|||||
 DB 551 EGIDQYRGWEN 561

RESULT 13
 AG3425
 l0E protein [imported] - *Brucella melitensis* (strain 16M)
 C/Species: *Brucella melitensis*
 C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
 C/Accession: AG3425
 R/DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A/Reference number: AD3252; PMID:11756688
 A/Accession: AG3425
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-350 <KUR>
 A/Cross-references: UNIPROT:O8YFX3; UNIPARC:UPI0000058056; GB:AE008917; PIDN:AAL52570.1;
 A/Experimental source: strain 16M
 C/Genetics:
 A/Gene: BMEI1389
 A/Map position: 1

Query Match 39.2%; Score 45.5; DB 2; Length 350;
 Best Local Similarity 47.4%; Pred. No. 28;
 Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNOSI-GNYRGWEN 19
 |||||:|||||
 DB 248 EFNPTGRQGVYGGYQGWV 266

RESULT 14
 AB2954
 conserved hypothetical protein Atu3235 [imported] - *Agrobacterium tumefaciens* (strain C58
 C/Species: *Agrobacterium tumefaciens*
 C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
 C/Accession: AB2954
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
 Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, I.
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AB2954
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-350 <KUR>
 A/Cross-references: UNIPROT:O8UAY2; UNIPARC:UPI00001647C7; GB:AE008689; PIDN:AAL44051.1;
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Gene: Atu3235
 A/Map position: linear chromosome

Query Match 39.2%; Score 45.5; DB 2; Length 350;
 Best Local Similarity 47.4%; Pred. No. 28;
 Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNOSI-GNYRGWEN 19
 |||||:|||||
 DB 248 EFNPTGRQGVYGGYQGWV 266

RESULT 15
 H98328
 hypothetical protein AGR_L_3161 [imported] - *Agrobacterium tumefaciens* (strain C58, Cerec
 C/Species: *Agrobacterium tumefaciens*
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C/Accession: H98328
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: H98328
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-370 <KUR>
 A/Cross-references: UNIPROT:O8UAY2; UNIPARC:UPI00000D211E; GB:AE007870; PIDN:AAK90154.1;
 C/Genetics:
 A/Gene: AGR_L_3161

A:Map position: linear chromosome

Query Match 39.2%; Score 45.5; DB 2; Length 370;
Best Local Similarity 47.4%; Pred. No. 29;
Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNQSI-GNYRGWFN 19
| | | | | : | | | |
Db 268 EFNPTGRQGVGGYQGVN 286

Search completed: February 15, 2006, 22:25:22
Job time : 40 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:17:44 ; Search time 228 Seconds
(without alignments)
61.888 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNQSIGNVRGWFNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_eprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	431	2 Q938V0_STRMU	Q938v0 streptococc
2	116	100.0	431	2 Q938V3_STRMU	Q938v3 streptococc
3	116	100.0	431	2 Q9AG98_STRMU	Q9ag98 streptococc
4	116	100.0	431	2 Q8DMW3_STRMU	Q8dmw3 streptococc
5	116	100.0	432	2 Q938V1_STRMU	Q938v1 streptococc
6	116	100.0	432	2 Q938V2_STRMU	Q938v2 streptococc
7	96	82.8	392	2 Q8MYI4_STRMR6	Q8myi4 streptococc
8	96	82.8	392	2 Q97N55_STRPN	Q97n55 streptococc
9	89	76.7	447	2 Q9AKA4_STRAG	Q9aka4 streptococc
10	89	76.7	447	2 Q8E2H1_STRAS	Q8e2h1 streptococc
11	89	76.7	447	2 Q8E7X9_STRP3	Q8e7x9 streptococc
12	84	72.4	169	2 Q8XTT4_STRAS	Q8xtt4 streptococc
13	84	72.4	169	2 Q8E3F4_STRP3	Q8e3f4 streptococc
14	79	68.1	129	2 Q5M137_STRT1	Q5m137 streptococc
15	79	68.1	461	2 Q56SA7_STRTR	Q56sa7 streptococc
16	79	68.1	474	2 Q5M6K4_STRT2	Q5m6k4 streptococc
17	79	68.1	482	2 Q5M5M6_STRT2	Q5m5m6 streptococc
18	79	68.1	485	2 Q5W212_STRT1	Q5w212 streptococc
19	77	66.4	398	2 Q5XEL1_STRP6	Q5xell1 streptococc
20	77	66.4	398	2 Q9A1Z8_STRPY	Q9a1z8 streptococc
21	77	66.4	398	2 Q7CNQ7_STRP8	Q7cnq7 streptococc
22	77	66.4	398	2 Q8P1I8_STRP3	Q8p1i8 streptococc
23	74	63.8	544	2 Q84OV8_STRMU	Q84ov8 streptococc
24	74	63.8	544	2 Q84OW6_STRMU	Q84ow6 streptococc
25	74	63.8	544	2 Q84OX3_STRMU	Q84ox3 streptococc
26	74	63.8	544	2 Q8DUR7_STRMU	Q8dur7 streptococc
27	70	60.3	211	2 Q8DVU8_STRIT	Q8dvu8 streptococc
28	70	60.3	226	2 Q93RG6_STRIT	Q93rg6 streptococc
29	55	47.4	332	2 Q8ZEM9_STRPE	Q8zem9 yersinia pe
30	55	47.4	337	2 Q8CL50_YERPE	Q8cl50 yersinia pe
31	53.5	46.1	906	2 Q55KP5_CRYNE	Q55kp5 cryptococcu

RESULT 1
Q938V0_STRMU
ID Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 0SD38D8DB8C4609F CRC64;

Query Match 100.0%; Score 116; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNVRGWFNP 20
|||||
Db 403 QEANYAGNQSIGNVRGWFNP 422

RESULT 2
Q938V3_STRMU
ID Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucan-binding protein B.

Qskat1 cryptococcus
Q82CS4 streptomycete
P38370 myxococcus
Q5en12 cryptospori
Q4xv24 plasmodium
Q4hu65 gibberella
Q619j3 aeromonas v
Q9rmh8 aeromonas h
Q9r9s7 aeromonas p
Q9lcj5 aeromonas p
Q6zic2 oryza sativ
Q5dd85 schistosoma
Q5ky63 geobacillus
Q8gex0 chromobacte

OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1] _
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=5J32;
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
 RT "Cloning of the gbpB gene from Streptococcus mutans."
 RL J. Dent. Res. 79:224-224 (2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=5J32;
 RX MEDLINE=21481971; PubMed=11598068;
 RA DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 RA Duncan M.J.;
 RT "Cloning of the Streptococcus mutans gene encoding glucan binding
 RT protein B and analysis of genetic diversity and protein production in
 RT clinical isolates."
 RL Infect. Immun. 69:6931-6941 (2001).
 DR EMBL; AF046410; AAK94500.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PRO1852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;
 Query Match 100.0%; Score 116; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0;
 QY 1 QYANYAGNQSIGNYRGWFP 20
 DB 403 QYANYAGNQSIGNYRGWFP 422

RESULT 3
 Q9AG98_STRMU
 ID Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
 AC Q9AG98
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
 GN Name=saga;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21153617; PubMed=11254612;
 RA DOI=10.1128/IAI.69.4.2493-2501.2001;
 RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
 RT "Identification of stress-responsive genes in Streptococcus mutans by
 RT differential display reverse transcription-PCR."
 RL Infect. Immun. 69:2493-2501 (2001).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21481977; PubMed=11598074;
 RA DOI=10.1128/IAI.69.11.6987-6998.2001;
 RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
 RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
 RT wall integrity and the maintenance of cell shape in Streptococcus
 RT mutans."
 RL Infect. Immun. 69:6987-6998 (2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;

RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=3VF4;
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
 RT "Cloning of the gbpB gene from Streptococcus mutans."
 RL J. Dent. Res. 79:224-224 (2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=3VF4;
 RX MEDLINE=21481971; PubMed=11598068;
 RA DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 RA Duncan M.J.;
 RT "Cloning of the Streptococcus mutans gene encoding glucan binding
 RT protein B and analysis of genetic diversity and protein production in
 RT clinical isolates."
 RL Infect. Immun. 69:6931-6941 (2001).
 DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.
 DR EMBL; AY046411; AAK94501.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PRO1852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 SQ SEQUENCE 431 AA; 44592 MW; 3BEE21FC5E47232E CRC64;
 Query Match 100.0%; Score 116; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0;
 QY 1 QYANYAGNQSIGNYRGWFP 20
 DB 403 QYANYAGNQSIGNYRGWFP 422

RESULT 4
 Q8DWM3_STRMU
 ID Q8DWM3_STRMU PRELIMINARY; PRT; 431 AA.
 AC Q8DWM3
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative secreted antigen GbpB/SagA; putative peptidoglycan
 DE hydrolase.
 GN Name=gbpB; OrderedLocusNames=SMU.22;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
 DR EMBL; AE014855; AAN57811.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PRO1852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 431 AA; 44620 MW; 2DICA685248CCD3E CRC64;
 Query Match 100.0%; Score 116; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QEANYAGNOSIGNYRGWFNP 20
    |||||
DB 403 QEANYAGNOSIGNYRGWFNP 422

RESULT 5
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1_
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-NAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AV046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 100.0%; Score 116; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QEANYAGNOSIGNYRGWFNP 20
    |||||
DB 404 QEANYAGNOSIGNYRGWFNP 423

RESULT 6
Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2_
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-NAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
```

```
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AV046413; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 100.0%; Score 116; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QEANYAGNOSIGNYRGWFNP 20
    |||||
DB 404 QEANYAGNOSIGNYRGWFNP 423

RESULT 7
Q8DMV4_STRR6
ID Q8DMV4_STRR6 PRELIMINARY; PRT; 392 AA.
AC Q8DMV4_
DT 01-NAR-2003 (TREMBlrel. 23, Created)
DT 01-NAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-NAR-2004 (TREMBlrel. 26, Last annotation update)
DE General stress protein GSP-781.
GN Name=gsp-781; OrderedLocusNames=spr2021;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544334;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren D.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008566; AAL00823.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 82.8%; Score 96; DB 2; Length 392;
Best Local Similarity 78.9%; Pred. No. 2.1e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 2 EANYAGNOSIGNYRGWFNP 20
    |:|||||:|||||
DB 360 ESNYAGNRTIGNHRCWFNP 378
```

```

RESULT 8
Q97N55 STRPN
ID Q97N55_STRPN PRELIMINARY; PRT; 392 AA.
AC Q97N55;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Secreted 45 kd protein.
GN Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Mayhew L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007509; AAK76264.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR TIGR; SP2216; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 82.8%; Score 96; DB 2; Length 392;
Best Local Similarity 78.9%; Pred. No. 2.1e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFNP 20
|:|||||:|||||
Db 360 ESNYAGNRTGNHGRGWFNP 378

RESULT 9
Q9AKA4 STRAG
ID Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE PcsB protein precursor.
GN Name=pcsb;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1311;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus."
RL J. Bacteriol. 183:1175-1183(2001).

```

```

DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Signal.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
Best Local Similarity 84.2%; Pred. No. 3.1e-05;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFNP 20
|:|||||:|||||
Db 417 ESNYAGNMSIGNRGSFNP 435

RESULT 10
Q8E2H1 STRAS
ID Q8E2H1_STRAS PRELIMINARY; PRT; 447 AA.
AC Q8E2H1;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tetelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wesels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014192; AAM98925.1; -; Genomic_DNA.
DR TIGR; SAG0017; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
Best Local Similarity 84.2%; Pred. No. 3.1e-05;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNRGWFNP 20
|:|||||:|||||
Db 417 ESNYAGNMSIGNRGSFNP 435

RESULT 11
Q8E7X9 STRA3
ID Q8E7X9_STRA3 PRELIMINARY; PRT; 447 AA.
AC Q8E7X9;
DT 01-MAR-2003 (TRENBLrel. 23, Created)

```

01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 Q8E3F4 STRA3 PRELIMINARY; PRT; 169 AA.
 Query Match 72.4%; Score 84; DB 2; Length 169;
 Best Local Similarity 73.7%; Pred. No. 6.4e-05;
 Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EANYAGNQSIGNYRGWFP 20
 DB 136 ESNYAGHQWIDNYRGWFP 154

RESULT 13
 Q8E3F4 STRA3 PRELIMINARY; PRT; 169 AA.
 ID Q8E3F4; AC Q8E3F4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein gbs1805.
 GN OrderedLocusNames=gbs1805;
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=216495;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=NEM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Meadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P., Kunst F.,
 RA "Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease."
 RL Mol. Microbiol. 45:1499-1513 (2002).
 DR EMBL; AL766843; CAD45661.1; -; Genomic_DNA.
 DR SgaList; gbs0016; -.
 DR InterPro; IPR007921; SgaP.
 DR InterPro; IPR009148; SgaP.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS0911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
 Best Local Similarity 84.2%; Pred. No. 3.1e-05;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EANYAGNQSIGNYRGWFP 20
 DB 417 ESNYAGNQSIGNYRGWFP 435

RESULT 12
 Q8DXT4 STRA5 PRELIMINARY; PRT; 169 AA.
 ID Q8DXT4 STRA5 PRELIMINARY; PRT; 169 AA.
 AC Q8DXT4; ID Q8DXT4; DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein SAG1762.
 GN OrderedLocusNames=SAG1762.
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=216466;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
 RA Tettelin H., Maignien V., Cieslewicz M.J., Eisen J.A., Peterson S.N., Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D., Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R., Radu D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S., Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M., Iacobini E.T., Brettoni C., Galli G., Martini M., Vegni F., Malone D., Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G., Fraser C.M.;
 RA "Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
 DR EMBL; AE014271; AAN00625.1; -; Genomic_DNA.
 DR TIGR; SAG1762; -.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; SgaP.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS0911; CHAP; 1.

RA Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcours J.,
 RA Goffeau A., Hols P.;
 RT "Complete sequence and comparative genome analysis of the dairy
 RT bacterium Streptococcus thermophilus.";
 RL Nat. Biotechnol. 22:1554-1558(2004).
 DR EMBL: CP000024; AAV62043.1; -; Genomic_DNA.
 DR InterPro: IPR007921; CHAP.
 DR Pfam: PF05257; CHAP; 1.
 DR PRINTS: PR01852; SIAPROTEIN.
 DR PROSITE: PS50911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 129 AA; 13136 MW; 7ADB47F41B7D4B3 CRC64;

Query Match 68.1%; Score 79; DB 2; Length 129;
 Best Local Similarity 73.7%; Pred. No. 0.0029;
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
 Db 97 ESNYAGNMLIGNYRGTFDP 115

RESULT 15

Q56SA7 STRTR PRELIMINARY; PRT; 461 AA.
 AC Q56SA7;
 DT 10-MAY-2005 (TREMBlrel. 30, Created)
 DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)
 DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)
 DE Cse.
 OS Streptococcus thermophilus.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1308;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CNR2365;
 RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
 RA Decaris B., Leblond-Bourget N.;
 RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
 RT Involved in Cellular Segregation in Streptococcus thermophilus.";
 RL J. Bacteriol. 187:2737-2746(2005).
 DR EMBL: AY695844; AAW33692.1; -; Genomic DNA.
 SQ SEQUENCE 461 AA; 48498 MW; CA72A95F945BD7A4 CRC64;

Query Match 68.1%; Score 79; DB 2; Length 461;
 Best Local Similarity 73.7%; Pred. No. 0.0012;
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
 Db 429 ESNYAGNMLIGNYRGTFDP 447

Search completed: February 15, 2006, 22:24:38
 Job time : 230 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:36:04 ; Search time 165 Seconds
(without alignments)
50.646 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNQSIGNYRGWFNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	20	4	US-10-383-930-22
2	116	100.0	20	5	US-10-797-821-22
3	116	100.0	431	4	US-10-383-930-29
4	116	100.0	431	4	US-10-383-930-30
5	116	100.0	431	4	US-10-383-930-33
6	116	100.0	431	5	US-10-797-821-29
7	116	100.0	431	5	US-10-797-821-30
8	116	100.0	431	5	US-10-797-821-33
9	116	100.0	432	4	US-10-383-930-31
10	116	100.0	432	4	US-10-383-930-32
11	116	100.0	432	5	US-10-797-821-31
12	116	100.0	432	5	US-10-797-821-32
13	96	82.8	392	5	US-10-472-928-4652
14	96	82.8	399	5	US-10-617-320-3230
15	77	66.4	398	5	US-10-474-792-600
16	52	44.8	188	4	US-10-156-761-12798
17	49	42.2	618	4	US-10-425-115-325660
18	47	40.5	198	4	US-10-425-115-304530
19	47	40.5	290	4	US-10-424-599-196081
20	47	40.5	509	4	US-10-437-963-105017
21	47	40.5	512	4	US-10-425-115-302311
22	47	40.5	524	4	US-10-425-114-68132
23	47	40.5	838	5	US-10-836-164-10
24	46	39.7	25	3	US-09-847-102A-107
25	46	39.7	25	3	US-09-847-102A-108
26	46	39.7	123	4	US-10-425-115-347949
27	46	39.7	186	5	US-10-450-763-55253

28	46	39.7	232	5	US-10-732-923-18914	Sequence 18914, A
29	46	39.7	278	4	US-10-029-386-32852	Sequence 32852, A
30	46	39.7	319	4	US-10-369-493-23683	Sequence 23683, A
31	46	39.7	319	4	US-10-282-122A-43119	Sequence 43119, A
32	46	39.7	440	4	US-10-094-749-1912	Sequence 1912, Ap
33	46	39.7	523	5	US-10-450-763-55252	Sequence 55252, A
34	46	39.7	546	4	US-10-112-944-712	Sequence 712, App
35	46	39.7	546	5	US-10-450-763-40380	Sequence 40380, A
36	46	39.7	599	3	US-09-847-102A-37	Sequence 37, Appl
37	46	39.7	682	4	US-10-152-548-16	Sequence 16, Appl
38	46	39.7	685	3	US-09-847-102A-57	Sequence 57, Appl
39	46	39.7	685	4	US-10-146-474-14	Sequence 14, Appl
40	46	39.7	685	4	US-10-301-764-14	Sequence 14, Appl
41	46	39.7	685	5	US-10-882-586A-9	Sequence 9, Appl
42	46	39.7	694	3	US-09-847-102A-56	Sequence 56, Appl
43	46	39.7	694	4	US-10-205-823-144	Sequence 144, App
44	46	39.7	694	4	US-10-285-976-53	Sequence 53, Appl
45	46	39.7	694	4	US-10-295-027-1344	Sequence 1344, Ap

ALIGNMENTS

RESULT 1

US-10-383-930-22
; Sequence 22, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-22

Query Match 100.0%; Score 116; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QEANYAGNQSIGNYRGWFNP 20

RESULT 2

US-10-797-821-22
; Sequence 22, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GdpB peptide
US-10-797-821-22

Query Match      100.0%; Score 116; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
Db 1 QEANYAGNQSIGNYRGWFNP 20

RESULT 3
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match      100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 4
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match      100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 5
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33

Query Match      100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 6
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match          100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
    |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 7
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match          100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
    |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 8
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
```

```
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match          100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
    |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422

RESULT 9
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

Query Match          100.0%; Score 116; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
    |||||
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 10
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
```

```
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match      100.0%; Score 116; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 11
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match      100.0%; Score 116; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 12
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
```

```
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match      100.0%; Score 116; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 13
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (O.E+01)
US-10-472-928-4652

Query Match      82.8%; Score 96; DB 5; Length 392;
Best Local Similarity 78.9%; Pred. No. 7.3e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EANYAGNQSIGNYRGWFNP 20
Db 360 ESNYAGNRTIGNHRGWFP 378

RESULT 14
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
```

Search completed: February 15, 2006, 22:39:15
Job time : 166 secs

```
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/617,320
;   FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
;   APPLICATION NUMBER: 60/ 085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 399 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: YES
;   ORIGINAL SOURCE:
;     ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...399
;     SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230
```

Query Match 82.8%; Score 96; DB 5; Length 399;
Best Local Similarity 78.9%; Pred. No. 7.4e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 EANYAGNOSIGNRGWFNP 20
      |:|||||:|||||:|||||
Db      367 ESNYAGNRTIGNHRCWFNP 385
```

```
RESULT 15
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
;   APPLICANT: Olmsted, Stephen
;   APPLICANT: Zagursky, Robert
;   APPLICANT: Nickbarg, Elliot
;   APPLICANT: Winter, Lourie
;   TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
;   FILE REFERENCE: AM 100399
;   CURRENT APPLICATION NUMBER: US/10/474,792
;   CURRENT FILING DATE: 2003-10-14
;   NUMBER OF SEQ ID NOS: 674
;   SOFTWARE: PatentIn version 3.0
;   SEQ ID NO 600
;   LENGTH: 398
;   TYPE: PRT
;   ORGANISM: Streptococcus pyogenes
US-10-474-792-600
```

Query Match 66.4%; Score 77; DB 5; Length 398;
Best Local Similarity 63.2%; Pred. No. 0.0049;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy      2 EANYAGNOSIGNRGWFNP 20
      |:|||||:|||||:|||||
Db      371 ESNYGRQYIADHRCWFNP 389
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:36:39 ; Search time 17 Seconds
(without alignments)
16.720 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNSIGNYRGWFNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_New.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	431	7	US-11-052-554A-210 Sequence 210, App
2	77	66.4	398	7	US-11-052-554A-252 Sequence 252, App
3	74	63.8	544	7	US-11-052-554A-358 Sequence 358, App
4	70	60.3	211	7	US-11-052-554A-352 Sequence 352, App
5	46	39.7	702	6	US-10-510-386-214 Sequence 214, App
6	42	36.2	106	7	US-11-004-590-390 Sequence 390, App
7	42	36.2	987	7	US-11-052-554A-153 Sequence 153, App
8	41	35.3	138	7	US-11-089-872-3 Sequence 3, Appli
9	41	35.3	213	7	US-11-158-505-34 Sequence 34, Appl
10	41	35.3	794	6	US-10-485-517-355 Sequence 355, App
11	40	34.5	108	6	US-10-771-257-39 Sequence 39, Appl
12	40	34.5	108	7	US-11-127-677-39 Sequence 39, Appl
13	40	34.5	3157	7	US-11-052-554A-142 Sequence 142, App
14	39.5	34.1	227	7	US-11-170-653-27 Sequence 27, Appl
15	39.5	34.1	3300	7	US-11-052-554A-133 Sequence 133, App
16	39	33.6	94	7	US-11-097-812-9 Sequence 9, Appli
17	39	33.6	106	6	US-10-485-517-181 Sequence 181, App
18	39	33.6	106	7	US-11-004-590-388 Sequence 388, App
19	39	33.6	107	7	US-11-064-174-173 Sequence 173, App
20	39	33.6	107	7	US-11-064-174-179 Sequence 179, App
21	39	33.6	107	7	US-11-154-337-5 Sequence 5, Appli
22	39	33.6	107	7	US-11-182-908-5 Sequence 5, Appli
23	39	33.6	108	6	US-10-665-658-3 Sequence 3, Appli
24	39	33.6	108	6	US-10-834-397-14 Sequence 14, Appl
25	39	33.6	108	7	US-11-120-338-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-11-052-554A-210
; Sequence 210, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 210
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-210

Query Match 100.0%; Score 116; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNSIGNYRGWFNP 20
|||||

Db 403 QEANYAGNSIGNYRGWFNP 422
|||||

RESULT 2

US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763

```
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match      66.4%; Score 77; DB 7; Length 398;
Best Local Similarity 63.2%; Pred. No. 0.00018;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
Db 371 ESNYSGROYIADHRGWFP 389

RESULT 3
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match      63.8%; Score 74; DB 7; Length 544;
Best Local Similarity 60.0%; Pred. No. 0.00068;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 QEANYAGNOSIGNYRGWFNP 20
Db 511 KEANYKNOQYISNFRGWFP 530

RESULT 4
US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match      60.3%; Score 70; DB 7; Length 211;
Best Local Similarity 57.9%; Pred. No. 0.00011;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
Db 178 ESNYKDOQWVDNYRGWFDP 196

RESULT 5
US-10-510-386-214
; Sequence 214, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 214
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-214

Query Match      39.7%; Score 46; DB 6; Length 702;
Best Local Similarity 43.5%; Pred. No. 15;
Matches 10; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 1 QEANYAGNOSIGNYR----GWFN 19
Db 435 QELSPAGGVKKGYRGNTIGWAN 457

RESULT 6
US-11-004-590-390
; Sequence 390, Application US/11004590
; Publication No. US20060008883A1
; GENERAL INFORMATION:
; APPLICANT: Lazar, Gregory Alan
; APPLICANT: Desjarlais, John R.
; TITLE OF INVENTION: METHODS OF GENERATING VARIANT PROTEINS WITH INCREASED HOST STRING
; FILE OF INVENTION: CONTENT AND COMPOSITIONS THEREOF
; FILE REFERENCE: 185832/US/5
; CURRENT APPLICATION NUMBER: US/11/004,590
; CURRENT FILING DATE: 2004-12-03
; PRIOR APPLICATION NUMBER: US 60/527,167
; PRIOR FILING DATE: 2003-12-04
; PRIOR APPLICATION NUMBER: US 60/581,613
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 60/601,665
; PRIOR FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: US 60/619,483
; PRIOR FILING DATE: 2004-10-14
; NUMBER OF SEQ ID NOS: 458
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 390
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-004-590-390

Query Match      36.2%; Score 42; DB 7; Length 106;
Best Local Similarity 54.5%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 NQSIGNYRGWF 18
Db 18 NQSIGNYRGWF 18
```


Db 26 SQNNYLGWY 36

RESULT 7

US-11-052-554A-153

; Sequence 153, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 153

; LENGTH: 987

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis H37RV

US-11-052-554A-153

Query Match 36.2%; Score 42; DB 7; Length 987;

Best Local Similarity 42.9%; Pred. No. 82;

Matches 9; Conservative 0; Mismatches 8; Indels 4; Gaps 1;

QY 3 ANYAGNQSIGN----YRGWFN 19

Db 901 AGIGGSGFGSGSLSGWFN 921

RESULT 8

US-11-089-872-3

; Sequence 3, Application US/11089872

; Publication No. US20050260212A1

; GENERAL INFORMATION:

; APPLICANT: Zhang, Yongke

; APPLICANT: Ramakrishnan, Vanitha

; APPLICANT: Law, Debbie

; TITLE OF INVENTION: ANTI-LFL2 ANTIBODIES FOR THE DIAGNOSIS, PROGNOSIS AND TREATMENT

; TITLE OF INVENTION: OF CANCER

; FILE REFERENCE: 05882.0223 NPUS02

; CURRENT APPLICATION NUMBER: US/11/089,872

; CURRENT FILING DATE: 2005-03-25

; PRIOR APPLICATION NUMBER: 60/557,440

; PRIOR FILING DATE: 2004-03-26

; PRIOR APPLICATION NUMBER: 60/638,708

; PRIOR FILING DATE: 2004-12-22

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 3

; LENGTH: 138

; TYPE: PRT

; ORGANISM: Mus musculus

US-11-089-872-3

Query Match 35.3%; Score 41; DB 7; Length 138;

Best Local Similarity 85.7%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 GNYRGWF 18

Db 120 GNYRAWF 126

RESULT 9

US-11-158-505-34

; Sequence 34, Application US/11158505

; Publication No. US20060002921A1

; GENERAL INFORMATION:

; APPLICANT: WINSOR-HINES, DAWN

; APPLICANT: RAO, PATRICIA

; APPLICANT: RINGLER, DOUGLAS J

; APPLICANT: PONATH, PAUL

; TITLE OF INVENTION: OPTIMIZED DOSING OF ANTI-CD4 ANTIBODIES FOR TOLERANCE

; TITLE OF INVENTION: INDUCTION IN PRIMATES

; FILE REFERENCE: TLN-031

; CURRENT APPLICATION NUMBER: US/11/158,505

; CURRENT FILING DATE: 2005-06-21

; PRIOR APPLICATION NUMBER: 60/582,181

; PRIOR FILING DATE: 2004-06-22

; NUMBER OF SEQ ID NOS: 76

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 34

; LENGTH: 213

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Humanized CD8

; OTHER INFORMATION: antibody light chain construct

US-11-158-505-34

Query Match 35.3%; Score 41; DB 7; Length 213;

Best Local Similarity 50.0%; Pred. No. 26;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 GNQSIGNYRGWF 18

Db 25 GSQDINNYLAWY 36

RESULT 10

US-10-485-517-355

; Sequence 355, Application US/10485517

; Publication No. US20050256299A1

; GENERAL INFORMATION:

; APPLICANT: University of Sheffield

; APPLICANT: Biosynexus Incorporated

; APPLICANT: Foster, Simon

; APPLICANT: Mond, James

; TITLE OF INVENTION: Antigenic Polypeptides

; FILE REFERENCE: PI00629WO

; CURRENT APPLICATION NUMBER: US/10/485,517

; CURRENT FILING DATE: 2004-02-02

; PRIOR APPLICATION NUMBER: GB 0118825.9

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: GB 0200349.9

; PRIOR FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 424

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 355

; LENGTH: 794

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-10-485-517-355

Query Match 35.3%; Score 41; DB 6; Length 794;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 YRGWFN 19

Db 562 YRGWFN 567

RESULT 11

US-10-771-257-39

; Sequence 39, Application US/10771257

; Publication No. US20050288864A1

; GENERAL INFORMATION:

; APPLICANT: Medical Research Council

; APPLICANT: SISSA - Scuola Superiore Internazionale di Studi Avanzati


```
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 133
; LENGTH: 3300
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-133
```

```
Query Match      34.1%; Score 39.5; DB 7; Length 3300;
Best Local Similarity 43.5%; Pred. No. 6.4e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 7; Gaps 1;
```

```
Qy      4 NYAGNQSIGNVR-----GWFN 19
Db      544 NILGNANIGNYILGSGNVGDFN 566
```

```
Search completed: February 15, 2006, 22:39:38
Job time : 18 secs
```

2006 FEB 16 08:44:51

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.754 Seconds
(without alignments)
1668.609 Million cell updates/sec

Title: US-10-797-821-31

Perfect score: 2110

Sequence: 1 MKKRILSAVLNVTLLSAT.....SIGNVRGWNFGSVSYIPN 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_21.*
1: Geneseq1990s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2110	100.0	432	7 ADD93651	Add93651 Streptoco
2	2110	100.0	432	9 ADX37274	Adx37274 Streptoco
3	2099	99.5	432	7 ADD93652	Add93652 Streptoco
4	2099	99.5	432	9 ADX37275	Adx37275 Streptoco
5	2083.5	98.7	431	7 ADD93650	Add93650 Streptoco
6	2083.5	98.7	431	9 ADX37273	Adx37273 Streptoco
7	2079.5	98.6	431	9 ADD93649	Add93649 Streptoco
8	2079.5	98.6	431	9 ADX37272	Adx37272 Streptoco
9	2079.5	98.6	431	9 AEB91500	Aeb91500 Microbial
10	2075.5	98.4	431	7 ADD93653	Add93653 Streptoco
11	2075.5	98.4	431	9 ADX37276	Adx37276 Streptoco
12	1234.5	58.5	447	5 ABP29684	Abp29684 Streptoco
13	1234.5	58.5	447	8 ADU69524	Adu69524 S agalact
14	1234.5	58.5	447	8 ADV88392	Adv88392 Streptoco
15	1234.5	58.5	447	8 ADV79645	Adv79645 Streptoco
16	1234.5	58.5	447	8 AEB91542	Aeb91542 Microbial
17	1093	51.8	398	9 AEB91542	Aeb91542 Streptoco
18	1087	51.5	398	5 ABP25919	Abp25919 Streptoco
19	1087	51.5	398	8 ADR38894	Adr38894 S. pyogen
20	1077.5	51.1	395	5 ABP25918	Abp25918 Streptoco
21	1059	50.2	392	6 ABU02747	Abu02747 S. pneumo
22	1059	50.2	392	8 ADT50226	Adt50226 S. pneumo
23	1059	50.2	392	8 ADT50165	Adt50165 S. pneumo
24	1059	50.2	399	8 ADR94595	Adr94595 Novel S.

ALIGNMENTS

RESULT 1

ADD93651

ID ADD93651 standard; protein; 432 AA.

XX AC ADD93651;

XX XX

DT 29-JAN-2004 (first entry)

XX XX

DE Streptococcus mutans glucan binding protein-B.

XX XX

KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX XX (FORS-) FORSYTH INST.

PA Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

DR GENBANK; AY046412.

XX Composition useful as vaccines for dental caries comprises a fragment of

PT a glucan binding protein-B binding to a major histocompatibility complex

XX class II protein.

PS Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding

CC protein-B (GbpB) of Streptococcus mutans strain 15Jp2. The sequence

CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The

CC invention provides immunogenic compositions and vaccines for dental

CC caries. The compositions comprise major histocompatibility complex (MHC)

CC class II protein-binding GbpB peptides, especially HLA-binding peptides,

CC covalently linked with peptide subunits of a glucosyltransferase. The

CC compositions are used in a claimed method of eliciting production of an

CC antibody in a mammal. Dieptopic or multiptopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies

CC raised against MHC class II binding fragments of GbpB can be used in

Aea58465 Streptoco
Adk47859 Streptoco
Adt50227 S.pneumon
Aar14150 MSP encod
Aar14530 Usp45 pro
Abb55584 Lactococc
Abu29734 Protein e
Adc95468 E. faeciu
Aay22579 Bacterial
Adv16553 E. faecal
Adh88105 Enterococ
Aay00250 Enterococ
Abp43469 E faecali
Abu88497 E. faecal
Abu13748 Enterococ
Ady16734 E. faecal
Ady39216 Novel Ent
Adh87829 Enterococ
Aay00251 Enterococ
Abp43470 E faecali
Abu88498 E. faecal

25 1059 50.2 399 9 AEA58465
26 1058 50.1 392 8 ADK47859
27 1058 50.1 392 8 ADT50227
28 681.5 32.3 461 2 AAR14150
29 680.5 32.3 461 2 AAR14530
30 660 31.3 456 5 ABB55584
31 588.5 27.9 524 6 ABU29734
32 584.5 27.7 525 7 ADC95468
33 561 26.6 210 2 AAY22579
34 529 25.1 482 9 ADV16553
35 529 25.1 497 7 ADH88105
36 500.5 23.7 449 2 AAY00250
37 500.5 23.7 449 5 ABP43469
38 500.5 23.7 449 6 ABU88497
39 500.5 23.7 449 9 ABU13748
40 500.5 23.7 449 9 ADV16734
41 500.5 23.7 449 9 ADV39216
42 500.5 23.7 450 7 ADH87829
43 452.5 21.4 422 2 AAY00251
44 452.5 21.4 422 5 ABP43470
45 452.5 21.4 422 6 ABU88498

CC	passive immunisation.	
XX		
SQ	Sequence 432 AA;	
	Query Match 100.0%; Score 2110; DB 7; Length 432;	
	Best Local Similarity 100.0%; Pred. No. 5.9e-134;	
	Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKKRILSAVLVSGVTLSSATTLSAIKADDFDAQIASQDSKINNLTAAQAAAAQAVNTIIG 60	
DB	1 MKKRILSAVLVSGVTLSSATTLSAIKADDFDAQIASQDSKINNLTAAQAAAAQAVNTIIG 60	
QY	61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120	
DB	61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120	
QY	121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAAANQE 180	
DB	121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAAANQE 180	
QY	181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAE 240	
DB	181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAE 240	
QY	241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTSDSAAQAVNN 300	
DB	241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTSDSAAQAVNN 300	
QY	301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGOWA 360	
DB	301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGOWA 360	
QY	361 ASAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVQGGQIQVQEBANYAGNOSIGNYRGW 420	
DB	361 ASAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVQGGQIQVQEBANYAGNOSIGNYRGW 420	
QY	421 FNPGSVSYIYPN 432	
DB	421 FNPGSVSYIYPN 432	
	RESULT 3	
	AD93652	
	ID AD93652 standard; protein; 432 AA.	
XX		
AC	AD93652;	
DT	29-JAN-2004 (first entry)	
XX		
DE	Streptococcus mutans glucan binding protein-B.	
XX		
KW	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.	
OS	Streptococcus mutans.	
XX		
PN	WO2003075845-A2.	
XX		
PD	18-SEP-2003.	
	Query Match 100.0%; Score 2110; DB 9; Length 432;	
	Best Local Similarity 100.0%; Pred. No. 5.9e-134;	
	Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKKRILSAVLVSGVTLSSATTLSAIKADDFDAQIASQDSKINNLTAAQAAAAQAVNTIIG 60	
DB	1 MKKRILSAVLVSGVTLSSATTLSAIKADDFDAQIASQDSKINNLTAAQAAAAQAVNTIIG 60	
QY	61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120	
DB	61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120	
QY	121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAAANQE 180	
DB	121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAAANQE 180	
QY	181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAE 240	
DB	181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAE 240	
QY	241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTSDSAAQAVNN 300	
DB	241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTSDSAAQAVNN 300	
QY	301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGOWA 360	
DB	301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGOWA 360	
QY	361 ASAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVQGGQIQVQEBANYAGNOSIGNYRGW 420	
DB	361 ASAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVQGGQIQVQEBANYAGNOSIGNYRGW 420	
QY	421 FNPGSVSYIYPN 432	
DB	421 FNPGSVSYIYPN 432	
	RESULT 2	
	ADX37274	
ID	ADX37274 standard; protein; 432 AA.	
XX		
AC	ADX37274;	
XX		
DT	21-APR-2005 (first entry)	
XX		
DE	Streptococcus mutant glucan binding protein B variant #3.	
XX		
KW	immunogenicity; immune stimulation; glucan binding protein-B;	
KW	microparticle; major histocompatibility complex; tooth disease.	
XX		
OS	Streptococcus mutans.	
XX		
PN	US2005031633-A1.	
XX		
PD	10-FEB-2005.	
XX		
PF	09-MAR-2004; 2004US-00797821.	
XX		
PR	13-APR-1998; 98US-0081550P.	
PR	08-JAN-1999; 99US-0115142P.	
PR	12-APR-1999; 99US-00290049.	
PR	07-MAR-2002; 2002US-0363209P.	
PR	08-AUG-2002; 2002US-0402483P.	
PR	07-MAR-2003; 2003US-00383930.	
XX		
PA	(SMIT/) SMITH D J.	
PA	(TAUB/) TAUBMAN M A.	
XX		
PI	Smith DJ, Taubman MA;	

XX 07-MAR-2003; 2003WO-US006962.
XX
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX
XX Smith DJ, Taubman MA;
PI WPI: 2003-845091/78.
DR GENBANK; AY046413.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 5; Page 8; 49pp; English.
XX
XX The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 3SNI. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
XX Sequence 432 AA;

Query Match 99.5%; Score 2099; DB 7; Length 432;
Best Local Similarity 99.3%; Pred. No. 3.2e-133;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTLTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSTLTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSKIVARNESLKQQAARSQAQVNTIOG 120
DB 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSKIVARNESLKQQAARSQAQVNTIOG 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEXMLQOEODKAAVEQKHQENQAANTVAANOE 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEXMLQOEODKAAVEQKHQENQAANTVAANOE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAKAAAEAAARQAQAQVNTIOG 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAKAAAEAAARQAQAQVNTIOG 240
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVWNGGOWA 360
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVWNGGOWA 360
QY 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGQGIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGQGIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 432

RESULT 4
ADX37275

ID ADX37275 standard; protein; 432 AA.
XX
AC ADX37275;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #4.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PP 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 3; SEQ ID NO 32; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 432 AA;

Query Match 99.5%; Score 2099; DB 9; Length 432;
Best Local Similarity 99.3%; Pred. No. 3.2e-133;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTLTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSTLTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSKIVARNESLKQQAARSQAQVNTIOG 120
DB 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSKIVARNESLKQQAARSQAQVNTIOG 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEXMLQOEODKAAVEQKHQENQAANTVAANOE 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEXMLQOEODKAAVEQKHQENQAANTVAANOE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAKAAAEAAARQAQAQVNTIOG 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAKAAAEAAARQAQAQVNTIOG 240
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300

QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCCTWGVKSLAPWVNTWNGGOWA 360
Db |||||
301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCCTWGVKSLAPWVNTWNGGOWA 360
QY 361 ASAAAAGYRVGTPSAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRCW 420
Db |||||
361 ASAAAAGYRVGTPSAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRCW 420
QY 421 FNPGSVSYIYPN 432
Db |||||
421 FNPGSVSYIYPN 432
RESULT 5
ADD93650
ID ADD93650 standard; protein; 431 AA.
XX AC
AC ADD93650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PP 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
DR GENBANK; AY046411.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 5; Page 8; 49pp; English.
XX
CC The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 3VF4. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Dieptopic or multieptopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
SQ Sequence 431 AA;
Query Match 98.7%; Score 2083.5; DB 7; Length 431;
Best Local Similarity 99.1%; Pred. No. 3.6e-132;
Matches 428; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
QY 1 MKKRILSVLVSGVTLSSTATTLSA KADDFDAQIASQDSKINNLTAAQQAAQVNTIQG 60
Db |||||
1 MKKRILSVLVSGVTLSSTATTLSA KADDFDAQIASQDSKINNLTAAQQAAQVNTIQG 60
QY 61 QVSALQTOQAEIQAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSAQKSNAAATSY 120

Db |||||
61 QVSALQTOQAEIQAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVBEKQKQENQAAINTVAANOE 180
Db |||||
121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVBEKQKQENQAAINTVAANOE 180
QY 181 TTAQNTNALNTQOAELEAAQLNLQAEELTTAQQKATLVAQKAAAEFAARQAAAAQAAAA 240
Db |||||
181 TTAQNTNALNTQOAELEAAQLNLQAEELTTAQQKATLVAQKAAAEFAARQAAAAQAAAA 240
QY 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQVNN 300
Db |||||
241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQVNN 299
QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCCTWGVKSLAPWVNTWNGGOWA 360
Db |||||
300 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCCTWGVKSLAPWVNTWNGGOWA 359
QY 361 ASAAAAGYRVGTPSAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRCW 420
Db |||||
360 ASAAAAGYRVGTPSAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRCW 419
QY 421 FNPGSVSYIYPN 432
Db |||||
420 FNPGSVSYIYPN 431
RESULT 6
ADX37273
ID ADX37273 standard; protein; 431 AA.
XX AC
AC ADX37273;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #2.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 3; SEQ ID NO 30; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II

QY 241 KAAAEKALQEQAAQAAAAAANNNTQATDASDQAAAAADNTQAAOTGSTDOSAAQAVNN 300
DB |||||
241 KAAAEKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAOTGSTDOSAAQAVNN 299
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGOCWTGWKSLAPWVGNVWNGGGOWA 360
DB |||||
300 SDOESTTATEAOPSSASSASTAVVTANTSSANTYPAGOCWTGWKSLAPWVGNVWNGGGOWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEBYAGNOSIGNYRGW 420
DB |||||
360 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEBYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYPN 432
DB |||||
420 FNPGSVSIYPN 431

RESULT 11
AD37276
ID AD37276 standard; protein; 431 AA.
AC
XX
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #5.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
OS Streptococcus mutans.

PN US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 33; 73pp; English.

PS The invention relates to a composition comprising a fragment of a glucan
XX binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
XX Streptococcus mutans GbpB protein of the invention.
XX Sequence 431 AA;

Query Match 98.4%; Score 2075.5; DB 9; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-131;
Matches 426; Conservative 3; Mismatches 3; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSATKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60

DB |||||
1 MKKRILSAVLVSGVTLSATKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
QY 61 QVSALQTOQAELOAENQRLQEAQSATLGQOIQTLSKIVARNESLKQARSQAKSNAATSY 120
DB |||||
61 QVSALQTOQAELOAENQRLQEAQSATLGQOIQTLSKIVARNESLKQARSQAKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDAAVEQKQOENQAAINTVAANOE 180
DB |||||
121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDAAVEQKQOENQAAINTVAANOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLAELTTAODQKATLVAQKAAAEAEAFQAAAAQAAAA 240
DB |||||
181 TIAQNTNALNTQQAQLEAAQLNLAELTTAODQKATLVAQKAAAEAEAFQAAAAQAAAA 240
QY 241 KAAAEKALQEQAAQAAAAAANNNTQATDASDQAAAAADNTQAAOTGSTDOSAAQAVNN 300
DB |||||
241 KAAAEKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAOTGSTDOSAAQAVNN 299
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGOCWTGWKSLAPWVGNVWNGGGOWA 360
DB |||||
300 SDOESTTATEAOPSSASSASTAVVTANTSSANTYPAGOCWTGWKSLAPWVGNVWNGGGOWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEBYAGNOSIGNYRGW 420
DB |||||
360 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEBYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYPN 432
DB |||||
420 FNPGSVSIYPN 431

RESULT 12
ABP29684
ID ABP29684 standard; protein; 447 AA.
XX
AC ABP29684;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 8544.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN70315.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
PS Claim 1; Page 3965; 4525pp; English.
XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 447 AA;

Query Match 58.5%; Score 1234.5; DB 5; Length 447;
Best Local Similarity 59.7%; Pred. No. 6.7e-75;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;
QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADFDQAIASQDSKINNLTAAQQAQAQVNTIQQ 60
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSQQAQAQVNTAIKG 58
QY 61 QVSNLQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQSAQKSNATSY 120
Db 59 QVGALESQSELEAQAQLEAVSQQLGQEIQTLSKNKIVARNESLKQVRSQAQKGN-LTNY 117
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQEQDKAAVEQKQOENQAINTVAANQE 180
Db 118 INTILNSKSVDAINRVSAIREVVSANERKMLQOQEQDKAAVEQKQOENQAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLAELTAAQLMQLAELTTAQDQKATLVQAQAAAEFAAQAQAQAQAAEA 240
Db 178 ALENKAALATQRAQLAQLAQLSALQTLTVQNEKASLIQAKAQAEAEARAKAAEAQAQAAEA 237
QY 241 KAAAEKALQEAQA 300
Db 238 KAQA 285
QY 301 SDQ--ESTTATA-----AQPSASSASTAA-----VAANTSSANTYPAG 336
Db 286 SSATTVAITTTATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPNG 345
QY 337 QCTWGVKSLAPVWGNVWNGGQWAAASAAAAGYRVGSTPSAGAVAVV--NDGGYGHVAVYT 394
Db 346 QCTWGAKSMAVWGNVWNGNQGASARAAGYSVGTTPRVGAVAVWPFYDGGYGHVAVVT 405
QY 395 GV-QGGQIQVQEAQVAGNOSIGNYRGWNP---GSVSYIYPN 432
Db 406 SVANNSSIQVMSNYAGNWSIGNYRGFNPSPASGSVYIYPN 447

RESULT 13
ADU69524
ID ADU69524 standard; protein; 447 AA.
XX
AC ADU69524;
XX
DT 10-FEB-2005 (first entry)
XX
DE S agalactiae hyperimmune serum reactive antigen seqid 219.
XX
KW immune stimulation; antigen; bacterial surface display;
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;
KW antibacterial; infection.
XX
OS Streptococcus agalactiae.
XX

PN WO2004099242-A2.
XX 18-NOV-2004.
XX 06-MAY-2004; 2004WO-EP004856.
XX 07-MAY-2003; 2003EP-00450112.
PR 28-NOV-2003; 2003EP-00450266.
XX (INTE-) INTERCELL AG.
XX
PI Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;
XX WPI; 2004-821662/81.
DR N-PSDB; ADU69307.
XX
PT New nucleic acid molecule encoding a hyperimmune serum reactive antigen,
PT useful for the manufacture of a vaccine against Streptococcus agalactiae
PT infection.
XX
PS Claim 14; SEQ ID NO 219; 22ipp; English.
XX
CC The invention describes an isolated nucleic acid molecule encoding a
CC hyperimmune serum reactive antigen or its fragment. Also described are: a
CC vector comprising the nucleic acid molecule; a host cell comprising the
CC vector; a hyperimmune serum-reactive antigen comprising a sequence
CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467
CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens
CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;
CC a process for producing a Streptococcus agalactiae hyperimmune serum
CC reactive antigen; a process for producing a cell that expresses a S.
CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical
CC composition, especially a vaccine, comprising the hyperimmune serum-
CC reactive antigen or nucleic acid molecule; an antibody that binds at
CC least to a selective part of the hyperimmune serum-reactive antigen; a
CC hybridoma cell line, which produces the antibody; a method for producing
CC the antibody; an antagonist that binds to the hyperimmune serum-reactive
CC antigen; a method for identifying an antagonist capable of binding to the
CC hyperimmune serum-reactive antigen; a method for identifying an
CC antagonist capable of reducing or inhibiting the interaction activity of
CC a hyperimmune serum-reactive antigen to its interaction partner; a
CC process for in vitro diagnosing a disease related to expression of the
CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis
CC of a bacterial infection, especially a S. agalactiae infection. The
CC hyperimmune serum reactive antigen is useful for isolating, purifying
CC and/or identifying an interaction partner of the hyperimmune serum
CC reactive antigen. The hyperimmune serum reactive antigen is useful for
CC generating a peptide binding to the hyperimmune serum reactive antigen,
CC where the peptide comprises anticalines, or for the manufacture of a
CC functional nucleic acid comprising aptamers or spiegelmers. The nucleic
CC acid molecule is useful for the manufacture of a functional ribonucleic
CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic
CC acid molecule hyperimmune serum-reactive antigen or antibody is useful
CC for the manufacture of a vaccine against S. agalactiae infection. This is
CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum
CC reactive antigen.
XX
SQ Sequence 447 AA;

Query Match 58.5%; Score 1234.5; DB 8; Length 447;
Best Local Similarity 59.7%; Pred. No. 6.7e-75;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;
QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADFDQAIASQDSKINNLTAAQQAQAQVNTIQQ 60
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSQQAQAQVNTAIKG 58
QY 61 QVSNLQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQSAQKSNATSY 120
Db 59 QVGALESQSELEAQAQLEAVSQQLGQEIQTLSKNKIVARNESLKQVRSQAQKGN-LTNY 117
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQEQDKAAVEQKQOENQAINTVAANQE 180
XX
SQ Sequence 447 AA;

Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALAEKQIENQNAINTVAANKQ 177
Qy 181 TIAQNTNALNTQQAQLEAAQLNLQALTTAQDQKATLVAKAAABEAARQAAAAQAAAEA 240
Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAAAEA 237
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGDSTDSQAQAVNN 300
Db 238 KAQAQAEAKQAESVAKAQAAA-----QVESATAPTETVOTQPRTEIKPSNLTAT 285
Qy 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336
Db 286 SSATTVAITATTATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMG 345
Qy 337 QCTGWKSLAPWVGNYGNGQWAAASAAAAGYRVGSTPSAGAVAVW--NDGGYGHVAVYT 394
Db 346 QCTWGAKSMAWVGNYGNGWNAQWGAASARAAGYSGVGTTPRVGAVAVWVPYDGGYGHVAVYT 405
Qy 395 GV-QGGQIQOQEANYAGNQSIGNYRGWNP---GSVSYIYPN 432
Db 406 SVANNSSIQVMESYAGNMSIGNYRGSPNPSAGSVYIYPN 447

RESULT 14

ID ADV88392 standard; protein; 447 AA.
AC ADV88392;
DT 24-FEB-2005 (first entry)
DE Streptococcus agalactiae protein sequence, SEQ ID 786.
XX
XX Antibacterial; Vaccine; bacterial infection.
XX
OS Streptococcus agalactiae.
XX
XX FR2824074-A1.
XX
PD 31-OCT-2002.
XX
XX 26-APR-2001; 2001PR-00005642.
PF
XX 26-APR-2001; 2001PR-00005642.
PR
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalloui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst P;
PI WPI; 2004-101891/11.
DR

Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

Claim 6; SEQ ID NO 786; 2687pp; French.

The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II; ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. CC Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2 contains 6617 sequence whereas the present patent only contains 2344 sequences.
XX
SQ Sequence 447 AA;

Query Match 58.5%; Score 1234.5; DB 8; Length 447;
Best Local Similarity 59.7%; Pred No. 6.7e-75;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

Qy 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIIG 60
Db 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIAATDSVINTLSGQQAQAAQVNTAIKG 58
Qy 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQITLSSKIARNESLQQAARSQKSNAAATSY 120
Db 59 QVGALESQSELEAQAQNAQLEAVSQQLGQBIQITLSNKIARNESLKKQVRSQKGN-LTNY 117
Qy 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQOQODKAAVFEKOENQAAINTVAANOE 180
Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALAEKQIENQNAINTVAANKQ 177
Qy 181 TIAQNTNALNTQQAQLEAAQLNLQALTTAQDQKATLVAKAAABEAARQAAAAQAAAEA 240
Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAAAEA 237
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGDSTDSQAQAVNN 300
Db 238 KAQAQAEAKQAESVAKAQAAA-----QVESATAPTETVOTQPRTEIKPSNLTAT 285
Qy 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336
Db 286 SSATTVAITATTATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMG 345
Qy 337 QCTGWKSLAPWVGNYGNGQWAAASAAAAGYRVGSTPSAGAVAVW--NDGGYGHVAVYT 394
Db 346 QCTWGAKSMAWVGNYGNGWNAQWGAASARAAGYSGVGTTPRVGAVAVWVPYDGGYGHVAVYT 405
Qy 395 GV-QGGQIQOQEANYAGNQSIGNYRGWNP---GSVSYIYPN 432
Db 406 SVANNSSIQVMESYAGNMSIGNYRGSPNPSAGSVYIYPN 447

RESULT 15

ADV81808
ID ADV81808 standard; protein; 447 AA.
XX
XX ADV81808;
XX
XX 24-FEB-2005 (first entry)
DT
XX Streptococcus agalactiae protein, SEQ ID 2949.
DE
XX Antibacterial; vaccine; bacterial infection.
XX
OS Streptococcus agalactiae.
XX
XX WO200292818-A2.
PN
XX 21-NOV-2002.
PD
XX 26-APR-2002; 2002WO-IB003059.
PF
XX 26-APR-2001; 2001PR-00005642.
PR
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalloui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
DR

PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
PS
PS Claim 6; SEQ ID NO 2949; 439pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
XX contains only 2344 sequences.
XX Sequence 447 AA;

Query Match 58.5%; Score 1234.5; DB 8; Length 447;
Best Local Similarity 59.7%; Pred. No. 6.7e-75;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;
QY 1 MKKRILSAVLVSGVTLSSATTLSAIAKADFDQAIAQSQDSKINNLTAAQQAQAQVNTIQC 60
DB 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIATDSVINTLSGQQAQAQNVTAIKG 58
QY 61 QVSALQTOQAEIQAENORLEAQSATIGQIQITLSSKIVARNESLKQQAARSQAOKSNAATSY 120
DB 59 QVGALESQSELEAQAQNALEAVSQQLGQEIQTLSNKIVARNESLKQVRSQAOKGN-LTNY 117
QY 121 INAIINSKSVDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQEQENQAINTVAANOE 180
DB 118 INTILNSKSVDAVNRVVAIREVVSANERKMLAQEQEADKALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLEAAQLMLQAELETTAQDOKATLVQAQAAAEAAQAAQAQAEEA 240
DB 178 ATEENKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAQAEEAARKAQAQAQAEEA 237
QY 241 KAAAEAKALQEQAAQQAQAANNNTQATDASDQQAQAADNTQAQTDGSDTQQAQAQVNN 300
DB 238 KAAQAQAQAQESVAKQAQA--QVESATAPTETVQTPRTETKPSNLTAT 285
QY 301 SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSSANTYPAG 336
DB 286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSTTPRAVSKPVRSYDSSNTYPNG 345
QY 337 QCTWGVKSLAPWVGNVWNGGQWASAAAAGYRVGSGTSPSAGAVAVW--NDGGYGHVAVYT 394
DB 346 QCTWGAKSMAWVGNVWNGNAGSARAAGYSVGTTPRVGAVVWVPYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQEAQYAGNOSIGNYRWENP---GSVSYIYPN 432
DB 406 SVANNSSIQVMESNYAGNMSIGNYRGSFNPSPASGSVYIYPN 447

Search completed: February 15, 2006, 18:02:59
Job time : 117.754 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:03:21 ; Search time 21.0162 Seconds
(without alignments)
1977.789 Million cell updates/sec

Title: US-10-797-821-31

Perfect score: 2110
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFMFGSVSIYVN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: Piri:.*
2: Piri2:.*
3: Piri3:.*
4: Piri4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	50.2	392	2 G95258	secreted 45 kd pro
2	1059	50.2	392	2 B98124	general stress pro
3	683.5	32.4	461	2 JN0097	secreted 45K prote
4	660	31.3	456	2 E89903	hypothetical prote
5	543.5	25.8	507	2 S05542	hypothetical prote
6	364	17.3	398	2 AC1763	peptidoglycan lyti
7	350.5	16.6	401	2 A11387	cell wall-binding
8	321	15.2	473	2 F70031	conserved hypotet
9	266.5	12.6	581	2 E75383	cell wall binding
10	260.5	12.3	436	2 AH1387	cell wall-binding
11	257	12.2	437	2 AB1763	cell wall-binding
12	256	12.1	461	2 H84099	conserved hypotet
13	242.5	11.5	528	2 B75310	tola protein - Esc
14	233	11.0	421	2 JV0057	membrane spanning
15	229.5	10.9	394	2 F90725	membrane spanning
16	229.5	10.9	394	2 G85576	probable tail fibe
17	229	10.9	971	2 B90835	probable membrane
18	229	10.9	973	2 C85693	cell wall-associat
19	227.5	10.8	392	2 F96937	hypothetical prote
20	226.5	10.7	255	2 G90061	hypothetical prote
21	226.5	10.7	265	2 B89837	hypothetical prote
22	226.5	10.7	267	2 F90028	hypothetical prote
23	226	10.7	492	2 A28616	M5 protein precurs
24	225.5	10.7	166	2 C90029	hypothetical prote
25	224.5	10.6	688	2 A83179	conserved hypotet
26	222	10.5	1122	2 G64887	probable tail fibe
27	219	10.4	4776	2 E95206	cell wall surface
28	218.5	10.4	1528	2 A60338	surface antigen A
29	216	10.2	483	2 A26297	M6 protein - Strep

RESULT 1

G95258
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: G95258
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95258
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-392 <KUR>
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:A5005672; F
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2216

ALIGNMENTS

30	213.5	10.1	536	2	D84325	Htr17 transducer [
31	213	10.1	2481	2	D90011	FmtB protein limpo
32	211	10.0	539	2	A28549	M24 protein precu
33	210.5	10.0	1566	2	A43607	cell surface antig
34	210.5	10.0	1773	2	A81937	IGA-specific metal
35	210	10.0	1589	2	T113606	hypothetical prote
36	208.5	9.9	436	2	S30284	M protein precurs
37	208	9.9	1561	2	S61314	IGA-specific metal
38	206.5	9.8	564	2	A60115	M protein precurs
39	206.5	9.8	1156	2	T34852	probable secreted
40	206	9.8	1081	2	S66736	transcription acti
41	204	9.7	2055	2	T31110	extracellular matr
42	203.5	9.6	351	2	B97273	uncharacterized pr
43	203.5	9.6	1815	2	C81169	IGA-specific metal
44	203	9.6	555	2	A26639	regulatory protein
45	202.5	9.6	574	2	S01272	regulatory protein

Db 315 TSAAAGFRTGTPQVGAICWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 374

Qy 420 WFPN-----GSVSYIY 430
||| | | | | |
Db 375 WFNPTTSEGFTYIY 390

RESULT 2

B98124
general stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: B98124
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: B98124
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-392 <KUR>
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317;
C:Genetics:
A:Gene: gsp-781

Query Match 50.2%; Score 1059; DB 2; Length 392;
Best Local Similarity 50.0%; Pred. No. 1.4e-44;
Matches 218; Conservative 75; Mismatches 91; Indels 52; Gaps 4;

Qy 1 MKKRILSAVLVSGVTLSATTLSAIFKADDDFAQIASQDSKINNLTAAQOAAQAVNTTQ 60
Db 1 MKKKILASILLSTVMVSQVAVLTAAETDDKIAAQDNKISNLTAAQOEAQKQVDQIQE 60

Qy 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQKSAQNAATSY 120
Db 61 QVSAIQASQSNLQANDRLQAESKLEGEITELSKNIVSRNQSLEKQARSQATNGAVTSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDKAAVEQKQENQAANTVAANQE 180
Db 121 INTIVNSKSITEAISRAVAMSEIVSANNKMLQEQKADKAISEKQVANNDAINTVIANQ 180

Qy 181 TIAQNTNALNTQOALEAAQLNLQAEITTAQDQKATLVAQKAAAEAAQAAAAQAAAEA 240
Db 181 KLADDAQALTTTQAEKAAELSLAAEKATAGEKASLLEQKAAAEAAEAFAVAEAAAYKE 240

Qy 241 KAAAEAKALQEAQAAQAAANNNTQATDASDQOAAAADNTQAAQTGSDTQSAQAQVNN 300
Db 241 KRAQQQSV-----LASANTNLTAQVQAQVSESAAA PRAKVRPT----- 279

Qy 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWGVNYWNGGGOWA 360
Db 280 -----YSTNASSYPIGECTWGVKTLAPWAGDYGWGAQWA 314

Qy 361 ASAAAAGRVGTPSAGAVAVNDGGYGHVAVVTGVQG-QQIQVQEAANYAGNQSTGNVRG 419
Db 315 TSAAAGFRTGTPQVGAICWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 374

Qy 420 WFPN-----GSVSYIY 430
||| | | | | |
Db 375 WFNPTTSEGFTYIY 390

RESULT 3

JN0097
secreted 45K protein precursor - Lactococcus lactis
C:Species: Lactococcus lactis
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: JN0097
R:van Asseldonk, M.; Rutten, G.; Oteman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.
Gene 95, 155-160, 1990

A:Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis sul
A:Reference number: JN0097; PMID:91071599; PMID:2123812
A:Accession: JN0097
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <VAN>
A:Cross-references: UNIPARC:UPI000016D72B; GB:M35374

Query Match 32.4%; Score 683.5; DB 2; Length 461;
Best Local Similarity 35.3%; Pred. No. 2.3e-26;
Matches 170; Conservative 107; Mismatches 135; Indels 69; Gaps 14;

Qy 1 MKKRILSAVLVSGVTLSATTLSAIFKADDDFAQIASQDSKINNLTAAQOAAQAVNTTQ 60
Db 1 MKKKIISAILMSTWILSAAAPLSGVYADT-NSDIADQDATISSAQSAKAQAQVDSLSQ 59

Qy 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQKSAQNAATSY 120
Db 60 KVDLSLQKQKQTSKQAIKIESEAKALNAQIATLNESEIKERTKTTLEAQAARSQVNSSATNY 119

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDKAAVEQKQENQAANTVAANQE 180
Db 120 MDVVNSKSLTDVIQKVTAIATVSSANKMLEQEQEKEQKSETVKKNYNQFVSLSQ 179

Qy 181 TIAQNTNALNTQOALEAAQLNLQAEITTAQDQKATLVAQKAAAEAAQAAAAQAAAEA 240
Db 180 SLDSQAQELTSQAQELKVATILNYQATIAQDQKQALLDEKAAAEKAAQAAKQAAYEA 239

Qy 241 KAAAEAKALQEAQAAQAAANNNTQATDASDQOAAAADNTQAAQTGSDTQ-----SAA 295
Db 240 Q-----QKEAAQAQAASTAATAKAVEAA-TSSASASSQAPOVSTSTDTNTTSNASAS 290

Qy 296 QAVNNSDOESTTATAAOPSASSASTAAVAANTSS-----ANTY 333
Db 291 NSSNSSNSSSSSSSSSSSSSSSSNSNAGGNTSGTSTGNTGGTTGGSGINSPIGNPY 350

Qy 334 PAGQCT---WG-----VKSLAPWGVNYWNGGGOWAASAAAAAG--YRVGTPSAGAVA 380
Db 351 AGGCTDVTWQYFAAQGIYIIRNIP-----GNGQWASNGPAGQVHLHVGAAP--GVIA 402

Qy 381 V-----WNDGGYGHVAVVTGVQG-QQIQVQEAANYAGNISIGNYRNFNPGSVSYIYP 431
Db 403 SFSADPFGVYANSPYGHVAIVKSVNSDGTITIKEGGY-GTTWGWHER-TVSASGVTFLLMP 460

Qy 432 N 432
|
Db 461 N 461

RESULT 4

E86903
hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86903
R:Bolotin, A.; Winkler, P.; Mager, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlic
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
A:Reference number: A86625; PMID:21235186; PMID:11337471
A:Accession: E86903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <STO>
A:Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI00000C6BFC; GB:AE005176; PID:g12725296; P1
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: usp45

Query Match 31.3%; Score 660; DB 2; Length 456;
Best Local Similarity 34.1%; Pred. No. 3.1e-25;
Matches 163; Conservative 111; Mismatches 136; Indels 68; Gaps 14;

Qy 1 MKKRILSAVLVSGVTLSATTLSAIFKADDDFAQIASQDSKINNLTAAQOAAQAVNTTQ 60

Db	1	MKKKIISATLMSTVLSAAPLSGVYADT-NSDIAKQDATISSAQAKAQAAQVDSLOS	59
Qy	61	QVSALQTQAELEAENORLEAQSATLGGQIOTLSSKIVARNESLKKQARSAAKSNAAATSY	120
Db	60	KVDSLQOKQASAKYAKIAKIESEKALNAQIATLINESIAERTKITLEAQARSQVNSSATNY	119
Qy	121	INAIINSKVSDAINRVSAIREVVSANERKMLQQOEQDKAAVEKQKQENQAAINTVAANOE	180
Db	120	MDAVNNSKSLTDVIQKVTAIATVSSANKQMLEQOEKEKELSKSETVKNTYKNTNQFVLSLQ	179
Qy	181	TTAQTNTALNTQQAOLEAQAQLNLOAELTTAODOKATLVAQKAAAEAAQAAAAQAAAAA	240
Db	180	SLDSQAQLTSQAELKVLATNYQATIAQDKKQSLDEKAAEAKAAQEAQKQAAYEA	239
Qy	241	KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDDTDOSSAQ--AV	298
Db	240	Q-----QKEAKAQAATAATTKAV--BEATSVSSSQASQSSSSNTSSNNTSS	288
Qy	299	NNSDQESTTATAAQPSASSASTAAVAANTSS-----ANTYPAG	336
Db	289	NSSSSSNSSSSSSSSSNGGGSTNTGNNAAGTNGTGGSSSGINSTPIANPYAGG	348
Qy	337	QCT--WG-----VKSLAPVGVNGVNGGWAASAAAAG--YRVGSTPAGAVAV--	381
Db	349	GCTDYVWYFAAQAAGIYRIMP-----GNGGQWATNGPAQGVHLHYVGAAP--GVIASSF	400
Qy	382	-----WNDGGYGHVAVYVGV--QGQGIQVQEAANYAGNQSIGNYRGMFNGSVSYIYPN	432
Db	401	SADPVGYANSPIGHVAIVSVNSNGTITKEGGY-GTTWGHGR-TVSASGVTFILMPN	456
RESULT 5			
S05542			
hypothetical protein, 54K - Enterococcus faecium			
C:Species: Enterococcus faecium			
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004			
C:Accession: S05542			
R:Puerst, P.; Moesch, H.U.; Solioz, M.			
Nucleic Acids Res. 17, 6724, 1989			
A:Title: A protein of unusual composition from Enterococcus faecium.			
A:Reference number: S05542; PMID:89385998; PMID:2780297			
A:Molecule type: DNA			
A:Residues: 1-507 <FUE>			
A:Cross-references: UNIPROT:PI3692; UNIPARC:UPI000016F6FC; GB:X16421; EMBL:M26048; NID:9			
A:Note: the authors translated the codon CGT for residues 221 and 223 as Lys			
Query Match 25.8%; Score 543.5; DB 2; Length 507;			
Best Local Similarity 28.9%; Pred. No. 1.5e-19;			
Matches 147; Conservative 90; Mismatches 170; Indels 101; Gaps 12;			
Qy	10	LVSGVTLSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAQAAQVNTTQGVVSALQTQQ	69
Db	1	MUSSTALTAVGPFAAABDFSQIQQQDKKIADLQNQQASQSQTEALEGGVSAINTKA	60
Qy	70	AELOAENORLEAQSATLGGQIOTLSSKIVARNESLKKQARSAAKSNAAATSYINAINSKS	129
Db	61	QDLTKQDTLRKESALQKEIKDLQERIEKREATIQKQARETVQKNTSSNYIDAVLNADS	120
Qy	130	VSDAINRVSAIREVVSANERKMLQQOEQDKAAVEKQKQENQAAINTVAANOETIAQNTNAL	189
Db	121	LADAVGRIQAMSTIVKANQDLVQOQKEDKQAEAKKAENEAKQKELADNQAALLESQKGLD	180
Qy	190	NTQQAOLEAAQLNLOAELTTAODOKATLVAQKAAAEAAQAAAAQAAAAA	235
Db	181	LAKQADNLNLTSLAAEQATAEDEKADLNKKAEEAEQARIREQARLAEARQQAQAEK	240
Qy	236	AAAAEAKAAEAKALQEQAAQAAAAANNNTQATDASDQAAA-----AAEEAARQAAAAQ	282
Db	241	AEKEAREQAAQAQAAQTAQALSSASTTTESSSAAQSSEESKAPESSTTESTESTSTTTE	300
Qy	283	AAQTGDDTDQSA----AQAVNNSDQESTTATAAQPSASSASTAAVAANTSANT-----	332

Db	301	NSSTGSSSTESSSTBEESTVPESSTQESTPANTSSSSSNTNVNNNTNNNTNNSTNNSTT	360
Qy	333	-----YPAGQCTWGVKSLA-----PWGVNMGWG-----	356
Db	361	NNNNNNNTVTPAPTPTPTPAPAPAPNPSPGVSUNGAAIVAEAYKYIGTPLYV---WGGKDPG	417
Qy	357	-----GOWAASAAAAAGYRVG-STPSAGAVAVVND--GGYGHVAVYVVG	395
Db	418	FDCSGFTRVYVLQVGTGRDIGMTVPQESAGTKISVQAKAGDLLFWGSAGGTVHVVAISLG	477
Qy	396	VOGGQIQVQEAANYAGNQSIGNYRGWFNP	423
Db	478	--GGQ-YIHAPQGENVKVGSVQ-WYTP	501
RESULT 6			
AC1763			
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clip11262)			
C:Species: Listeria innocua			
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004			
C:Accession: AC1763			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.			
Science 294, 849-852, 2001			
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, C.			
A:Title: Comparative genomics of Listeria species.			
A:Reference number: AB1077; PMID:21337279; PMID:11679669			
A:Accession: AC1763			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-398 <GLA>			
A:Cross-references: UNIPROT:Q927Y8; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:CAC97875.1;			
A:Experimental source: strain Clip11262			
C:Genetics:			
A:Gene: spl			
Query Match 17.3%; Score 364; DB 2; Length 398;			
Best Local Similarity 27.4%; Pred. No. 5.4e-11;			
Matches 115; Conservative 88; Mismatches 145; Indels 72; Gaps 16;			
Qy	1	MKKRILSAV-LVSGVTLSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAQAAQVNTTQ	59
Db	1	MKXNTFIAISLAAVISLTPAFTTNVFA--DVNTDIQNQDKKINDIKSKTKGLQSLSSLV	58
Qy	60	GOVSALQTQQAOLEAENORLEAQSATLGGQIOTLSSKIVARNESLKKQARSAAKSNAAAT	119
Db	59	ADLEKAQEKAKSLQEGFPQTGKELQNLQDIKDINERIKERETVLKERARAWQKTSNSNA	118
Qy	120	YINAIINSKVSDAINRVSAIREVVSANERKMLQQOEQD-----KAAVEQKQENQAAI	172
Db	119	YLEVILDAENLSDLVGRVSAVNVQLVDSKSLIEDQKDEKALKTKQTAVKKQEEQATAI	178
Qy	173	NTVAANOETIAQNTNALNTQQAOLEAQAQLNLOAELTTAODOKATLVAQKAAAEAAQAA	232
Db	179	HEFEAQQ-----NKIEAQAKAEKAIVAQLAADQASAEKAGLVSE--DKAAKE--	226
Qy	233	AAQAAAEKAAAEKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDDTQ	292
Db	227	-----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTS	258
Qy	293	SAAQAVNNSDQESTTATAAQPSASSASTAAVAANTSANTYPAG-----OCTWGVKSL	345
Db	259	NKVESTNNSEAFS-----AATPSGGGYSAMIAAARALQKPYLSLGATGPSAFPDCS-GFTSY	313
Qy	346	---APWVGNYWNGWGQWAAASAAAAGYRVGSTPSAGAVAVN--DGGYGHVAVYVVGQGI	401
Db	314	AFRAAGVSLPRTSGGGYAAAAASKIS----ASQAKPGDLVFFNYGGGIAHVGIYVG--GGQM	367
RESULT 7			

Best Local Similarity 24.8%; Pred. No. 4.le-06;
Matches 129; Conservative 68; Mismatches 174; Indels 149; Gaps 19;

Qy 16 LSSATTLSAIAKADDFDAQTASQ-----DSKINNLTAAQQAAAQAQVNTIOQG 61
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 40 LAGGDTLTSRRLEQLRLLIEQRRLSVQXKQLTEVRARLRQLNAQRAALDRDLATSE 99
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 62 VSALTQQAELQENORLEAQSYTLGO---OIQTSLSKIVARNESLKQOAARSQAQS--N 115
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 100 VTDLENELADVLA--RVTAATRALLRETEAQRIVTRSQV----EALKVDARAAMKALYRA 152
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 116 AATSINAIITINSKSVDAINRVSAIREVVSANEMKLQ-----QEODKAAP 161
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 153 RNTQYMRLUSNSISDMILRLDYANNAGORNVEVMEO LRGAABELTTQQLRQROOSDAL 212
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 162 EOKQEQNQ-----AAINTVAANQETIAQNTNALNTQQAQ----- 195
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 213 RLQGEQQTCLAELRDRTFQDALAELQRSACQQAQAVAVETQAQALTAQTIDS LGNV 272
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 196 -----LEAAQLNLQAE-----LTTAQPOKATLV A-- 219
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 273 VRERTLEBERRRRLFEERRREREAARRIRAEQERAKKEARLARIRAEEKERKAREAEA 332
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 220 --OKAAABEAARAAAAQAABAAEAKAEAKALOE---QAQAQAQAAN-----NNNTQAT 268
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 333 ARQKALAE--ARAAAQKVAEAAEKARAEAKARAEQAARAAAQAQAAAQAKAEQAARAQAE 390
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 269 DA-----SDQAAAAAADNT-----QAAQTGDSTDQSAQAQVVNSDOESTT---ATAAQP 313
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 391 AAARAQAQAQAAAAAFAEAQAARAQAQAQAQAQKVREQNAVAREORDLTQOOQT AQOE 450
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 314 SASSASTAAVAANTSSANTYPAGCOTWGVKSLPWPVGNYMGNGGGOWAASAAAAGYRGST 373
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 451 KQLAAELAPLPAISSRDGLFPFLP---GCKVAAP---YGTSGAQWEVLSCASGCGRAVAA 502
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 374 PSAGA VA-----VNNDGGYGHVAYVTGVGGGIQV 403
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 503 QTGNVIASAYAALGWILLDHGSNVITYGYFLGQDTLVEV 542

RESULT 10
AH1387
cell wall binding proteins homolog lmo2504 [imported] - Listeria monocytogenes (strain E)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH1387
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
. D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dusserget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Kreft, J., Kunl, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehlend,
A.:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; PMID:21537279; PMID:11679669
A:Accession: AH1387
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-436 <GLA>
A:Cross-references: UNIPROT:Q8Y4E2; UNIPARC:UPT0000055221; GB:NC_003210; PIDN:CAD00582.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2504

Query Match 12.3%; Score 260.5; DB 2; Length 436;
Best Local Similarity 23.3%; Pred. No. 6e-06;
Matches 98; Conservative 86; Mismatches 160; Indels 77; Gaps 14;

Qy 6 LSAILVSG-VTLSSATTLGAI-----KADFDQIASQDSKINNLTAAQQAAAQAQV 55
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 11 LSLLIISAPLTGANHDTINDMKRQNEIQQKKSEIDKNIDSKSELNHLESARKDADEL 70
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 56 NTIQGV-----SALOTQQAELQHNRLEAQSATLGQOIOTLSSKIVARNESLKKQAPSRA 111
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

A:Gene: tola
A:Map position: 17 min
A:Start codon: GTG
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:14-34/Domain: transmembrane #status predicted <MSS>
F:78-301/Domain: helical #status predicted <HSP>
F:355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 11.0%; Score 233; DB 2; Length 421;
Best Local Similarity 28.5%; Pred. No. 0.00012;
Matches 117; Conservative 66; Mismatches 178; Indels 50; Gaps 19;

QY 5 ILSAVLVGVTLSSTLSSATLTAIADDPDAQIASQ-----DSKINNLTAAQQAQAQVNTIQ 59
DB 16 IISAVL--HVILFAALINSS-----FDENIEASAGGGSSIDAVMVDGAVVEQYKRMQ 68
QY 60 GOVSA-----LQTOQA--ELQ-----AENQRL--EAQSATLGGQIQTLSSKIVARNES 103
DB 69 SOESSAKRSDEQRKKEQQAELREKQQAELRKLQLEKRLAAQEQKKQAEAAKQAE 128
QY 104 LKQ-QARSQAQNAATSYINALINSKVSDAINRVSAIREVVSANERKMLQOQEQODKAAVE 162
DB 129 LKQQAEEAAAKAADAADAKAEAEAAK-----AAADAKKAEAEAAKAAAE 181
QY 163 -KQOENQAAI--NTVAANQETIAQNTNALNTQQAQLEAAQLNLQAL--TTAQDQKATLV 218
DB 182 AOKKAEEAAALKKKAEEAAAEAAAEARKKAATEAAEAEEKAAAEKAAAEKAAADKKA--A 239
QY 219 AOKAAAE--EAARQAAQAQAAAEAAKAAAEKALQEQAAQAAQAAANNNT--QATDASDQQA 275
DB 240 AEKAADKKAAAEAAKAAAEKAAAEKAAADKKAEEAAAEKAAAEKAAAEADDDIFG 299
QY 276 AAADNTQAAQTGSDTQSAQAQVNNSDQSTTATAAQ-PSASSASTAAVAANTSSANTYP 334
DB 300 ELSSGKNAPKTKGGAGNNASPAAGSNTKNGASGADINNYAQIKSAIESKFYDASSYA 359
QY 335 AQCTWGVKSLAP--WVGNYWNGGOWA-ASAAAAGYRVGSTPSAGAVAVW 382
DB 360 GKTCTLRK-LAPDGMMLDIKPEGDPALCOAALAAAKLAKIPKPPSQAVY 409

RESULT 15
F90725
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F90725
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90725
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <HAY>
A:Cross-references: UNIPROT:O8X965; UNIPARC:UPI0000D086A; GB:BA000007; PIDN:BA034197.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs0774

Query Match 10.9%; Score 229.5; DB 2; Length 394;
Best Local Similarity 27.5%; Pred. No. 0.00017;
Matches 109; Conservative 62; Mismatches 177; Indels 49; Gaps 15;

QY 5 ILSAVLVGVTLSSTLSSATLTAIADDPDAQIASQ-----DSKINNLTAAQQAQAQVNTIQ 59
DB 16 IISAVL--HVILFAALINSS-----FDENIEASAGGGSSIDAVMVDGAVVEQYKRMQ 68
QY 60 GOVSALQQAQLEAQENQLEAQATLGGQIQTLSSKIVARNESLQ-----QARSQAQKS 114
DB 69 SOESS-----AKRSDEQRKKEQQA-----BELREKQAEEQLKLEKRLAAEQKK 118

QY 115 NAATSYINAIINSKVSDAINRVSAIREVVSANERKMLQOQEQODKAAVEQKQEQENQAINT 174
DB 119 QAEEAAKQAEKQQAEEAAAKAADAADAKAEAEAAKAAADAADAKKAEAEAAKA 178
QY 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ--ABLTTAQDQKATLVAAQKAAAEAA--R 229
DB 179 AAEAOKKAEAAALKKKAEAAAEAAAEARKKAAAEKAAADKKA---AEKAAAEKAAADK 235
QY 230 QAAAAQAAAEAAKAAAEKALQEQAAQAAQAAANNNTQATDASDQQAQAAADNTQAAQTDGS 289
DB 236 KAAAEKAAADKKA--AKAAAEKAAAKAAA-----EADDIFGELSSGKQAPKTTGG 286
QY 290 TDQSAQAQVNNSDQSTTATAAQ-PSASSASTAAVAANTSSANTYPAGQCTWGVKSLAP- 347
DB 287 AKGNNASPAAGSNTKNGASGADINNYAQIKSAIESKFYDASSYAGKCTTLRIK-LAPD 345
QY 348 -WVGNYWNGGOWA-ASAAAAGYRVGSTPSAGAVAVW 382
DB 346 GMLLDIKPEGDPALCOAALAAAKLAKIPKPPSQAVY 382

Search completed: February 15, 2006, 18:10:23
Job time : 24.0162 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:21 ; Search time 123.762 Seconds
(without alignments)
2462.693 Million cell updates/sec

Title: US-10-797-821-31
Perfect score: 2110
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFMFGSVIYIPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result #	No.	Score	Query Match	Length	DB	ID	Description
1	2110	100.0	432	2	Q938V2	STRMU	Q938V2 streptococc
2	2099	99.5	432	2	Q938V1	STRMU	Q938V1 streptococc
3	2083.5	98.7	431	2	Q9AG98	STRMU	Q9AG98 streptococc
4	2079.5	98.6	431	2	Q938V3	STRMU	Q938V3 streptococc
5	2079.5	98.6	431	2	Q8DM3	STRMU	Q8DM3 streptococc
6	2075.5	98.4	431	2	Q938V0	STRMU	Q938V0 streptococc
7	1234.5	58.5	447	2	Q9AKA4	STRAG	Q9AKA4 streptococc
8	1234.5	58.5	447	2	Q8E2H1	STRAS	Q8E2H1 streptococc
9	1234.5	58.5	447	2	Q8E7X9	STRAS	Q8E7X9 streptococc
10	1093	51.8	398	2	Q5XEL1	STRP6	Q5XEL1 streptococc
11	1093	51.8	398	2	Q7CNO7	STRP8	Q7CNO7 streptococc
12	1093	51.8	398	2	Q8P318	STRP3	Q8P318 streptococc
13	1092.5	51.8	474	2	Q5M6K4	STRT2	Q5M6K4 streptococc
14	1092	51.8	485	2	Q5M212	STRT1	Q5M212 streptococc
15	1087	51.5	398	2	Q9A1Z8	STRPY	Q9A1Z8 streptococc
16	1059	50.2	392	2	Q8DMY4	STRR6	Q8DMY4 streptococc
17	1059	50.2	392	2	Q97N55	STRPN	Q97N55 streptococc
18	946	44.8	211	2	Q9ZAS7	STRMU	Q9ZAS7 streptococc
19	676.5	32.1	461	1	USP45	IACLC	P22867 lactococcus
20	660	31.3	456	2	Q9CDJ1	LACLA	Q9CDJ1 lactococcus
21	588.5	27.9	524	2	Q9K2J9	ENTFC	Q9K2J9 enterococcu
22	575.5	27.3	516	1	P54	ENTFC	P13692 enterococcu
23	549.5	26.0	576	2	Q9KJ3	ENTHR	Q9KJ3 enterococcu
24	549	26.0	482	2	Q5M5M6	STRT2	Q5M5M6 streptococc
25	544	25.8	470	2	Q8Q6E3	9LACT	Q8Q6E3 lactococcus
26	532	25.2	482	2	Q93LK4	ENTFA	Q93LK4 enterococcu
27	511.5	24.2	461	2	Q5GSA7	STRTR	Q5GSA7 streptococc
28	504.5	23.9	449	2	Q93LK3	ENTFA	Q93LK3 enterococcu
29	436.5	20.7	211	2	Q8DVU8	STRMU	Q8DVU8 streptococc
30	416	19.7	226	2	Q93RG6	STRIT	Q93RG6 streptococc
31	416	19.7	544	2	Q840W6	STRMU	Q840W6 streptococc

32	416	19.7	544	2	Q840X3	STRMU	Q840X3 streptococc
33	414	19.6	544	2	Q840V8	STRMU	Q840V8 streptococc
34	411	19.5	544	2	Q8DUR7	STRMU	Q8DUR7 streptococc
35	404.5	19.2	129	2	Q5M137	STRT1	Q5M137 streptococc
36	403	19.1	169	2	Q8E3F4	STRP3	Q8E3F4 streptococc
37	396	18.8	169	2	Q8DXT4	STRP5	Q8DXT4 streptococc
38	364	17.3	398	2	Q9Z7Y8	LISIN	Q9Z7Y8 listeria in
39	350.5	16.6	401	2	Q7IWS3	LISMF	Q7IWS3 listeria mo
40	350.5	16.0	401	2	Q9RE04	LISMO	Q9RE04 listeria mo
41	337.5	16.0	430	2	Q8LHV2	BACCR	Q8LHV2 bacillus ce
42	325	15.4	604	2	Q6M552	CORGL	Q6M552 corynebacte
43	324	15.4	427	2	Q73DG0	BACC1	Q73DG0 bacillus ce
44	321	15.2	473	1	YVCE	BACSU	P40767 bacillus su
45	319	15.1	600	2	Q8NQA0	CORGL	Q8NQA0 corynebacte

ALIGNMENTS

RESULT 1
Q938V2 STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2_ (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates."
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match	100.0%;	Score 2110;	DB 2;	Length 432;
Best Local Similarity	100.0%;	Pred. No. 2.7e-86;	Mismatches 0;	Indels 0;
Matches 432;	Conservative 0;	Gaps 0;		
QY	1	MKKRILSAVLVSGVTLSATLSA	KADDFDAQIASQDSKINNLT	AAQAAAAQVNTIIG 60
Db	1	MKKRILSAVLVSGVTLSATLSA	KADDFDAQIASQDSKINNLT	AAQAAAAQVNTIIG 60
QY	61	QVSALQTOAELOAENRLEAQS	ATLGQIQTLSSKIVARNESLKQ	ARSAAKSNAAATSY 120
Db	61	QVSALQTOAELOAENRLEAQS	ATLGQIQTLSSKIVARNESLKQ	ARSAAKSNAAATSY 120
QY	121	INAIINSKSVSDAINRVSAIR	EVVSAKMLQQQEQDQKAAV	EQKQENQAINTVAANQE 180
Db	121	INAIINSKSVSDAINRVSAIR	EVVSAKMLQQQEQDQKAAV	EQKQENQAINTVAANQE 180
QY	181	TTIAQNTNALNTQQAOLEAA	QAQLNLAQLTAQDQKATL	VAQKAAAEARQAAAAA 240

```
|||||
181 TIAQNTNALNTQQAQLAAQLNLQAEELTTAQQKATLVAKAAAEAAEAARQAAAAQAAREA 240
|||||
241 KAAAEKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDQSAQAQAVNN 300
|||||
241 KAAAEKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDQSAQAQAVNN 300
|||||
301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGNWNGGOWA 360
|||||
301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGNWNGGOWA 360
|||||
361 ASAAAAGYRVGTSFGAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEAANTAGNOSIGNYRGW 420
|||||
361 ASAAAAGYRVGTSFGAGAVAVWNDGGYGHVAVVTGVQGGQIQVQEAANTAGNOSIGNYRGW 420
|||||
421 FNPQSVSYIYPN 432
|||||
421 FNPQSVSYIYPN 432

RESULT 2
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AV046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 99.5%; Score 2099; DB 2; Length 432;
Best Local Similarity 99.3%; Pred. No. 8.3e-86;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRLTSLAVLGVGVTLSATLSSATLTKADDFDAQIASQDSKINNLTAQQAQAQVNTIQG 60
DB 1 MKKRLTSLAVLGVGVTLSATLSSATLUSAKDDFDAQIASQDSKINNLTAQQAQAQVNTIQG 60

QY 61 QVSALQTOQAELQAEQNRLQEAQATLGGQIQTLSSKIVARNESLKQARSQAQSAQNTSY 120
DB 61 QVSALQTOQAELQAEQNRLQEAQATLGGQIQTLSSKIVARNESLKQARSQAQSAQNTSY 120

QY 121 INAIINSVSDAINRVSAIRVSVANEKMLQQQODKAAVEQKQEQEAINTVAANQE 180
DB 121 INAIINSVSDAINRVSAIRVSVANEKMLHQEQDKAAVEQKQEQEAINTVAANQE 180
```

RESULT 3

```
Q9AG98_STRMU
ID Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
AC Q9AG98;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
GN Name=saga;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21153617; PubMed=11254612;
RX DOI=10.1128/IAI.69.4.2493-2501.2001;
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
RT "Identification of stress-responsive genes in Streptococcus mutans by
RT differential display reverse transcription-PCR.";
RN Infect. Immun. 69:2493-2501(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RX DOI=10.1128/IAI.69.11.6987-6998.2001;
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
RT wall integrity and the maintenance of cell shape in Streptococcus
RT mutans.";
RN Infect. Immun. 69:6987-6998(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
```



```

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014855; RAN57811.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;

Query Match      98.6%; Score 2079.5; DB 2; Length 431;
Best Local Similarity 98.8%; Pred. No. 6.1e-85;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTTTLSAIAKDDFDFAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTTTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQQEQDKAAVEQKQENQAAINTVAAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQQEQDKAAVEQKQENQAAINTVAAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAQAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAQAAEA 240

QY 241 KAAAEKAKLOEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300
DB 241 KAAAEKAKLOEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300

QY 301 SDQESTTTAAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360
DB 301 SDQESTTTAAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360

QY 361 ASAAAAGRVGTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420
DB 361 ASAAAAGRVGTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 6
Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RC Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RC STRAIN=5SM3;
RC MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8B8C4609F CRC64;

Query Match      98.4%; Score 2075.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 9.2e-85;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTTTLSAIAKDDFDFAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTTTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQQEQDKAAVEQKQENQAAINTVAAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQQEQDKAAVEQKQENQAAINTVAAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAQAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAQAAEA 240

QY 241 KAAAEKAKLOEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300
DB 241 KAAAEKAKLOEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300

QY 301 SDQESTTTAAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360
DB 301 SDQESTTTAAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360

QY 361 ASAAAAGRVGTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420
DB 361 ASAAAAGRVGTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RC MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183 (2001).
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```

```
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Signal.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.5%; Score 1234.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2.le-47;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTILGTAAV--TVNADDFDSKIATDSVINTLSGQAAAQNVTAIKG 58
QY 61 QVSALQTOQAELQAEQNRLEAQSATLGQIQIOTLSSKIVARNESLKQOASAKSNAATSY 120
DB 59 QVGALESQQSELEAQAQAEVSVQQLGQBIOTLSNKIVARNESLKKQVRSQKGN-LTNY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEXMLQOQODKAAVEQKQOENQAANTVAANO 180
DB 118 INTILNSKSVSDAVNRVVAIREVWSANEXMLQOQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAKAAAEAEAAQAAAAA 240
DB 178 ALENKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEEAARKAAEAQAAA 237
QY 241 KAAAEKALQEAQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300
DB 238 KAAEAQAQAESVAKAQAQA-----QVESATAPTETVQTPTEIKPSNLAT 285
QY 301 SDQ-----ESTTATA-----AQSASSASTAA-----VAANTSANTYPAG 336
DB 286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYSSNTYPNG 345
QY 337 QCTWGVKSLAPVWGNVGNWGQWAAASAAAAGYRVGSTPSAGAVAVM--NDGGYGHVAVYT 394
DB 346 QCTWGAKSMAWVGNVGNWGNWQASARAAGYSGVGTTPRVGAVAVWVPYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQEAQYAGNQSIGNYRGWFPN-----GSVSYIYPN 432
DB 406 SVANNSSIQVMESNAGNMSGNYSIGNYRGSPNPSAGSVYIYPN 447

RESULT 9
Q8E7X9 STRA3 PRELIMINARY; PRT; 447 AA.
AC Q8E7X9_
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsB; OrderedLocustNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carly H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae.";
Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
EMBL; AE014192; AM98925.1; -; Genomic_DNA.
TIGR; SAG0017; -.
InterPro; IPR007921; CHAP.
InterPro; IPR009148; Siba.
Pfam; PF05257; CHAP; 1.
PRINTS; PRO1852; SIBAPROTEIN.
PROSITE; PS0911; CHAP; 1.
Complete proteome.
KW SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.5%; Score 1234.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2.le-47;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTILGTAAV--TVNADDFDSKIATDSVINTLSGQAAAQNVTAIKG 58
QY 61 QVSALQTOQAELQAEQNRLEAQSATLGQIQIOTLSSKIVARNESLKQOASAKSNAATSY 120
DB 59 QVGALESQQSELEAQAQAEVSVQQLGQBIOTLSNKIVARNESLKKQVRSQKGN-LTNY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEXMLQOQODKAAVEQKQOENQAANTVAANO 180
DB 118 INTILNSKSVSDAVNRVVAIREVWSANEXMLQOQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAKAAAEAEAAQAAAAA 240
DB 178 ALENKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEEAARKAAEAQAAA 237
QY 241 KAAAEKALQEAQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300
DB 238 KAAEAQAQAESVAKAQAQA-----QVESATAPTETVQTPTEIKPSNLAT 285
QY 301 SDQ-----ESTTATA-----AQSASSASTAA-----VAANTSANTYPAG 336
DB 286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYSSNTYPNG 345
QY 337 QCTWGVKSLAPVWGNVGNWGQWAAASAAAAGYRVGSTPSAGAVAVM--NDGGYGHVAVYT 394
DB 346 QCTWGAKSMAWVGNVGNWGNWQASARAAGYSGVGTTPRVGAVAVWVPYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQEAQYAGNQSIGNYRGWFPN-----GSVSYIYPN 432
DB 406 SVANNSSIQVMESNAGNMSGNYSIGNYRGSPNPSAGSVYIYPN 447
```



```
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

  Query Match      51.8%; Score 1093; DB 2; Length 398;
  Best Local Similarity 53.8%; Pred. No. 3.5e-41;
  Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQG 60
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKKQARSQAQKSNAAATSY 120
DB 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQENQAAINTVAANQE 180
DB 118 INALLNSKISIDVNRVLNRAVNSANAKLLEQKADKVSLEEKQAANQTAINTAANWA 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQALAEITTAQDQKATLVAKAAAEBAARQAAQAAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALQEAQAQAAQAAANNNNTQATDASQQAADNTQAAQTGSDTQDSAAQAVNN 300
DB 238 KA-----QEQA-----QQAASVEAAKSAITPAQATPAQSSNA 272
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
DB 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSAPWAGNWNNGGQWA 325
QY 361 ASAAAAGYRVGSTPSAGAVAVNWDGGYGHVAVVTVGQ-CGQIQVOEANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGTPMVGAIVNWDGGYGHVAVVVEVQSASSIRVMEISYSGRQYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 12
QBP318_STRP3 PRELIMINARY; PRT; 398 AA.
AC QBP318; Q7CFL7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP80015, SpyM3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22133808; PubMed=12129345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12129345; DOI=10.1073/pnas.152298499;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus;
RT bacterium Streptococcus thermophilus.";
```

```
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL; BA000034; BAC63110.1; -; Genomic DNA.
DR EMBL; AF014136; AM78621.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

  Query Match      51.8%; Score 1093; DB 2; Length 398;
  Best Local Similarity 53.8%; Pred. No. 3.5e-41;
  Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQG 57
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKKQARSQAQKSNAAATSY 120
DB 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQENQAAINTVAANQE 180
DB 118 INALLNSKISIDVNRVLNRAVNSANAKLLEQKADKVSLEEKQAANQTAINTAANWA 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQALAEITTAQDQKATLVAKAAAEBAARQAAQAAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALQEAQAQAAQAAANNNNTQATDASQQAADNTQAAQTGSDTQDSAAQAVNN 300
DB 238 KA-----QEQA-----QQAASVEAAKSAITPAQATPAQSSNA 272
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
DB 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSAPWAGNWNNGGQWA 325
QY 361 ASAAAAGYRVGSTPSAGAVAVNWDGGYGHVAVVTVGQ-CGQIQVOEANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGTPMVGAIVNWDGGYGHVAVVVEVQSASSIRVMEISYSGRQYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 13
QSM6K4_STRT2 PRELIMINARY; PRT; 474 AA.
AC QSM6K4;
DT 01-FEB-2005 (TREMBlrel. 29, Created)
DT 01-FEB-2005 (TREMBlrel. 29, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Glucan binding protein (PcGB).
GN Name=PcGB; OrderedLocusNames=stu0022;
OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=264199;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kulakauskas S., Lapidus A., Goldsman E., Mazur M., Pusch G.D.,
RA Fongstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
```



```

RL Nat. Biotechnol. 22:1554-1558(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LMG 18311;
RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Leblond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus.";
RL J. Bacteriol. 187:2737-2746(2005).
DR EMBL; CP000023; AAV59752.1; -; Genomic DNA.
DR EMBL; AY730643; AAW82375.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BAC6A9E0A0A200D CRC64;

Query Match 51.8%; Score 1092.5; DB 2; Length 474;
Best Local Similarity 51.7%; Pred. No. 4.4e-41;
Matches 247; Conservative 73; Mismatches 107; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQ 60
DB 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYDQIAATNNAISNLASQQAQAQVATIQS 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQAARSQAQNTATSY 117

QY 121 INAIINSKVSDAINRVSAIREVSVANEKMLQQQODKAAVEKQOENQAANTVAANO 180
DB 118 INSILNSKISSEAITRITAISKVVTANNLLTKQSDQKELAAKQOENQAANTIAANKS 177

QY 181 TTAQNTNALNTQOALEAAQLNLOAELTTAQDOKATLVAQKAAAEAAARQAAA---AQA 236
DB 178 ELETTEAGLTQQAELEAAQVTLAAELATAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237

QY 237 AAEAKA-----AAEAKALQEAQAQAQAANNNTQATDASDQQAQAANTQAAQTG 287
DB 238 IAESSEATAQVVASSEAAATSVASVEAATSEAVAPATSEAPVATSEAPVAPVAPVAPV 296

QY 288 DSTDQSAQAQVNN-----SDQSTTATAQPS-----ASSASTAAV-----AANT- 327
DB 297 EVQPESSAAPVSEAPVAPVATSEAPATSEAPVATSEAPVATSEAPVAPVAPVAPVAPV 356

QY 328 -----SSANTYPAGCTGWKSLAPWGNVGNWGOWAASAAAAGVYRGSTPSAGAVAVW 382
DB 357 KVSAASTPNTYPVQCTGWKSLAPWGNVGNWGNKWNWLNASQAAGHSVGTTPVAGAIW 416

QY 383 -ND-GGYGHVAVYTVQVG-QGIVQVEANYAGNQSIGNYRGWFPN-----GSVSYIYP 431
DB 417 PNDGGYGHVAVYTVSAGSANSIQWESNYAGNMSISNYRGTFDPTSSAHGGSVFIYP 474

RESULT 14
QSM212_STR11
ID QSM212_STR11 PRELIMINARY; PRT; 485 AA.
AC QSM212;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN Name=pcdB; OrderedLocusNames=str0022;
OS Streptococcus thermophilus (strain CNR2 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=299768;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,

```

```

RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV61641.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 51.8%; Score 1092; DB 2; Length 485;
Best Local Similarity 50.9%; Pred. No. 4.8e-41;
Matches 249; Conservative 71; Mismatches 107; Indels 62; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQ 60
DB 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYDQIAATNNAISNLASQQAQAQVATIQS 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQAARSQAQNTATSY 117

QY 121 INAIINSKVSDAINRVSAIREVSVANEKMLQQQODKAAVEKQOENQAANTVAANO 180
DB 118 INSILNSKISSEAITRITAISKVVTANNLLTKQSDQKELAAKQOENQAANTIAANKS 177

QY 181 TTAQNTNALNTQOALEAAQLNLOAELTTAQDOKATLVAQKAAAEAAARQAAA---AQA 236
DB 178 ELETTEAGLTQQAELEAAQVTLAAELATAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237

QY 237 AAEAKA-----AAEAKALQEAQAQAQAANNNTQATDASDQQAQAANTQAAQTG 287
DB 238 IAESSEATAQVVASSEAAATSVASVEAATSEAVAPATSEAPVATSEAPVAPVAPVAPV 296

QY 288 DSTDQSAQAQVNN-----SDQSTTATA-----AOPSASSA----- 318
DB 297 EVQPESSAAPVSEAPVAPVATSEAPATSEAPVATSEAPVATSEAPVAPVAPVAPVAPV 356

QY 319 STAAVAANT-----SSANTYPAGCTGWKSLAPWGNVGNWGOWAASAAAAGVYRGV 371
DB 357 TSEAPAAETHKVSAASTPNTYPVQCTGWKSLAPWGNVGNWGNKWNWLNASQAAGHSV 416

QY 372 STPSAGAVAVW-ND-GGYGHVAVYTVQVG-QGIVQVEANYAGNQSIGNYRGWFPN----- 423
DB 417 TTPVAGAIWVNDGGYGHVAVYTVSAGSANSIQWESNYAGNMSISNYRGTFDPTSSAH 476

QY 424 -GSVSYIYP 431
DB 477 GGSVFIYP 485

RESULT 15
Q9A128_STRPY
ID Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; Q7BH59;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;

```


RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Pagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006474; AAK3158.1; -; Genomic DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome; signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;
Query Match 51.5%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.6%; Pred. No. 6.6e-41;
Matches 232; Conservative 58; Mismatches 107; Indels 36; Gaps 6;
QY 1 MKKRILSAVLVSGVTLSSNTLSAISKADDPDAQIASODSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSTKIAKQDSIIISNLTTEKAAQNVSAQQA 57
QY 61 QVSALQTOQAEIQAENQLEAQSATLGGQIQIOTLSSKIVARNESLKQQAASAKSNAATSY 120
DB 58 QVSSLSQSQDKLTARTLEALSKEFEQEIKALTSQIVARNEKLNQKQAESAYKNNETSGY 117
QY 121 INAIINSKSVDAINRVSAIREVVSANEKMLQOQODKAAVEKQKQENQAQNTVAANQE 180
DB 118 INALLNSKSIQDVNRLVAINRAVSAKALLBQKADKVSLEKQAAQNTAINTIAANMA 177
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTAQDQKATLVAKKAAEEAARQAAAAQAAEA 240
DB 178 MAEENQNTLRTOQANLVAAATANLALQLASATEDKANLVAKQEAKEKAAEAALAEQAQKV 237
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGDSTDQSAQAQVNN 300
DB 238 KA-----EQQA-----QQAASVEAKSAITPAQATPAQAQSSNA 272
QY 301 SDQESTTATAQPSASSASTAAVAANTSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
DB 273 IEPAALTAPAA-PSAGPQT-----SYDSSNTYFPVQCTWGAKSLAPWAGNWNWNGGQWA 325
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVVTGVQ-GGQIQVQEBANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGSTPMVGAIVAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 30.0232 Seconds
(without alignments)
1189.611 Million cell updates/sec

Title: US-10-797-821-31
Perfect score: 2110
Sequence: 1 MKRILSAVLVSGVTLSAT.....SIGNRGWNPQSGSVIYPN 432
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,AA:*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCUS_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/baCkfilesl.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	50.2	399	2	US-09-107-433-3230
2	1058	50.1	392	2	US-09-583-110-4374
3	681.5	32.3	461	1	US-08-186-222-2
4	584.5	27.7	525	2	US-09-107-532A-5095
5	561	26.6	210	2	US-09-222-938A-67
6	539	25.1	497	2	US-09-134-000C-5990
7	500.5	23.7	449	2	US-09-071-035-482
8	500.5	23.7	449	2	US-10-206-576-482
9	500.5	23.7	450	2	US-09-134-000C-5714
10	452.5	21.4	422	2	US-09-071-035-484
11	452.5	21.4	422	2	US-10-206-576-484
12	283.5	13.4	469	2	US-09-489-039A-13565
13	237	11.2	264	2	US-09-134-001C-5035
14	236	11.2	257	2	US-09-710-279-3244
15	236	11.2	267	2	US-09-134-001C-4539
16	235	11.1	477	2	US-09-302-540-11649
17	232	11.0	2310	2	US-09-874-923-120
18	226	10.7	1236	2	US-09-769-787-109
19	222.5	10.5	270	2	US-09-134-001C-5441
20	214.5	10.2	266	2	US-09-134-001C-5453
21	214.5	10.2	468	2	US-09-328-352-6321
22	214	10.1	149	2	US-09-710-279-1682
23	214	10.1	157	2	US-09-710-279-2870
24	213	10.1	440	2	US-08-302-756F-35
25	212.5	10.1	655	2	US-09-902-540-10005
26	212	10.0	610	2	US-09-336-447A-11
27	212	10.0	610	2	US-09-952-267B-11

28	208.5	9.9	1566	1	US-08-687-956A-23	Sequence 23, Appl
29	206	9.8	679	2	US-09-489-039A-12307	Sequence 12307, A
30	203.5	9.6	639	2	US-09-902-540-14908	Sequence 14908, A
31	203.5	9.6	1070	2	US-09-902-540-13861	Sequence 13861, A
32	202.5	9.6	955	1	US-08-006-676B-1	Sequence 1, Appl
33	202.5	9.6	955	1	US-08-282-845-2	Sequence 2, Appl
34	202.5	9.6	955	1	US-08-428-414A-3	Sequence 3, Appl
35	202.5	9.6	955	4	PCT-US94-00324-1	Sequence 1, Appl
36	197.5	9.4	1024	2	US-09-270-767-44973	Sequence 44973, A
37	196.5	9.3	883	2	US-09-489-039A-12755	Sequence 12755, A
38	195.5	9.3	443	1	US-08-795-475-6	Sequence 6, Appl
39	195.5	9.3	443	2	US-08-325-278B-6	Sequence 6, Appl
40	193	9.1	2107	2	US-09-949-016-7646	Sequence 7646, Ap
41	193	9.1	2107	2	US-09-949-016-7647	Sequence 7647, Ap
42	193	9.1	2115	2	US-09-296-662-33	Sequence 33, Appl
43	191.5	9.1	2101	1	US-08-466-390-4	Sequence 4, Appl
44	191.5	9.1	2101	1	US-08-470-950-4	Sequence 4, Appl
45	191.5	9.1	2101	1	US-08-467-781-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-107-433-3230
; Sequence 3230, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
City: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3230:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (8) LOCATION 1...399

SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

US-09-107-433-3230

Query Match 50.2%; Score 1059; DB 2; Length 399;
Best Local Similarity 50.0%; Pred. No. 3.6e-73;
Matches 218; Conservative 75; Mismatches 91; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIIKADDFDAQLASODSKINNLTAAQQAQAVNTIQQ 60
DB 8 MKKRILASLLLSSTVWMSQVAVUTTAETTTDKIAAQDNKI SNLTAAQQAQAVNTIQQ 67
QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIQTLSKIVARNESLKKQARSQAQNAATSY 120
DB 68 QVSAIQAESNLQAEENDRLQAESKKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 127
QY 121 INAIINSKVSDAINRVSAREVVSANERKMLQOQODKAAVEQKQOENQAAINTVAANO 180
DB 128 INTIVNSKSIETASIRVAAMSEIVSANNKMLEEQKADKKAISEKQVANNDAINTVIANQ 187
QY 181 TTAQNTNALNTQOAELEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240
DB 188 KLADDAQALTTKQAEKKAELSAAEKATAEGEKASLLEQKAAAEAEARAAVAEAAYKE 247
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300
DB 248 KRASQOQSV-----LASANTNLTAQVQAVSESAAAAPVRKVRPT----- 286
QY 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360
DB 287 -----YSTNASSYPIGECTGWGKTLAPWAGDYWNGGAQWA 321
QY 361 ASAAAAGYRVGTPSAGAVAVMNDGGYGHVAVVTGQVQ-GQIQVOEANYAGNOSIGNYRG 419
DB 322 TSSAAAAGFTGTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 381
QY 420 WFNPP-----GSVSYIY 430
DB 382 WFNPTTSEGFTYIY 397

RESULT 2
US-09-583-110-4374
; Sequence 4374, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4374
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4374

Query Match 50.1%; Score 1058; DB 2; Length 392;
Best Local Similarity 50.0%; Pred. No. 4.2e-73;
Matches 218; Conservative 74; Mismatches 92; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIIKADDFDAQLASODSKINNLTAAQQAQAVNTIQQ 60
DB 1 MKKRILASLLLSSTVWMSQVAVUTTAETTTDKIAAQDNKI SNLTAAQQAQAVNTIQQ 60
QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIQTLSKIVARNESLKKQARSQAQNAATSY 120

DB 61 QVSAIQAESNLQAEENDRLQAESKKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 120
QY 121 INAIINSKVSDAINRVSAREVVSANERKMLQOQODKAAVEQKQOENQAAINTVAANO 180
DB 121 INTIVNSKSIETASIRVAAMSEIVSANNKMLEEQKADKKAISEKQVANNDAINTVIANQ 180
QY 181 TTAQNTNALNTQOAELEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240
DB 181 KLADDAQALTTKQAEKKAELSAAEKATAEGEKASLLEQKAAAEAEARAAVAEAAYKE 240
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAAQAVNN 300
DB 241 KRASQOQSV-----LASANTNLTAQVQAVSESAAAAPVRKVRPT----- 279
QY 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360
DB 280 -----YSTNASSYPIGECTGWGKTLAPWAGDYWNGGAQWA 314
QY 361 ASAAAAGYRVGTPSAGAVAVMNDGGYGHVAVVTGQVQ-GQIQVOEANYAGNOSIGNYRG 419
DB 315 TSSAAAAGFTGTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 374
QY 420 WFNPP-----GSVSYIY 430
DB 375 WFNPTTSEGFTYIY 390

RESULT 3
US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007
; GENERAL INFORMATION:
; APPLICANT: Suri, Bruno
; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-186-222-2

Query Match 32.3%; Score 681.5; DB 1; Length 461;

```
Best Local Similarity 35.3%; Pred. No. 3.4e-44;
Matches 170; Conservative 106; Mismatches 136; Indels 69; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATLSSAIDDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSAIDDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQIOTLSSKIVARNESLKKQARSQAQKSNAAATSY 120
Db 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQIOTLSSKIVARNESLKKQARSQAQKSNAAATSY 120
QY 60 KVDLSLQKQKTSKAQIAKTESERKALNAQIATLINESIKERTKTLEAQAARSQAQVNSSATNY 119
Db 60 KVDLSLQKQKTSKAQIAKTESERKALNAQIATLINESIKERTKTLEAQAARSQAQVNSSATNY 119
QY 121 INAINSKSVSDAINRVSAIREVVSANEXKMLQOQEDKAQVQKQOENQAANTVAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVVSANEXKMLQOQEDKAQVQKQOENQAANTVAANQE 180
QY 120 MDVAVNSKSLTDVIOKVTATVSSANKQMLEQKEQKELQKSETVKKYNNQFVLSLQ 179
Db 120 MDVAVNSKSLTDVIOKVTATVSSANKQMLEQKEQKELQKSETVKKYNNQFVLSLQ 179
QY 181 TTAQNTNALNTQQAQLEAQLNLQAEELTTAQQKATLVQAQKAAAEAAQAAAAQAABA 240
Db 181 TTAQNTNALNTQQAQLEAQLNLQAEELTTAQQKATLVQAQKAAAEAAQAAAAQAABA 240
QY 180 SLDSQAQELTSQAQELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQKQAAYEA 239
Db 180 SLDSQAQELTSQAQELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQKQAAYEA 239
QY 241 KAAAEKALQEAQAQAQAANNNTQTADSDQAAAAAANTQAQGTGSDTQ-----SAA 295
Db 241 KAAAEKALQEAQAQAQAANNNTQTADSDQAAAAAANTQAQGTGSDTQ-----SAA 295
QY 240 Q-----QKEAAQAQAATAAKAVEAA-TSSASASSQAQVSTSTDTNTTSNASAS 290
Db 240 Q-----QKEAAQAQAATAAKAVEAA-TSSASASSQAQVSTSTDTNTTSNASAS 290
QY 296 QAVNSDQESTTATAAQPASASASTAAVANTSS-----ANTY 333
Db 296 QAVNSDQESTTATAAQPASASASTAAVANTSS-----ANTY 333
QY 291 NSSNSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 350
Db 291 NSSNSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 350
QY 334 PAQOCT-----WG-----VKSLAPVGVNGYNGGOWAAASAAAG--YRVGSTPSAGAVA 380
Db 334 PAQOCT-----WG-----VKSLAPVGVNGYNGGOWAAASAAAG--YRVGSTPSAGAVA 380
QY 351 AVGGCTDYQWYPAQAQIYIRIMP-----GNGGQWASNGPAQGVHLVVGGAAP--GVIA 402
Db 351 AVGGCTDYQWYPAQAQIYIRIMP-----GNGGQWASNGPAQGVHLVVGGAAP--GVIA 402
QY 381 V-----WPDGQYGHVAVTGVQV-GQIQVQEAANYAGNQIGNYRGWFGVSVSYIYP 431
Db 381 V-----WPDGQYGHVAVTGVQV-GQIQVQEAANYAGNQIGNYRGWFGVSVSYIYP 431
QY 403 SSFSADPVGYSNYPYGHVAVKSVNSDGTITIKGGY-GTTWGHGR-TVSASGVTFLLP 460
Db 403 SSFSADPVGYSNYPYGHVAVKSVNSDGTITIKGGY-GTTWGHGR-TVSASGVTFLLP 460
QY 432 N 432
Db 432 N 432
QY 461 N 461
Db 461 N 461

RESULT 4
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5095:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...525
; SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095

Query Match 27.7%; Score 584.5; DB 2; Length 525;
Best Local Similarity 30.4%; Pred. No. 1.1e-36;
Matches 160; Conservative 92; Mismatches 163; Indels 111; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATLSSAIDDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
Db 2 VKKRILSAVLVSGVTLSATLSSAIDDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 61
QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQIOTLSSKIVARNESLKKQARSQAQKSNAAATSY 120
Db 62 QVSEINTQAQDLAKQDTLRQSSAQLVKDIADLQERIEKREDTIQKQAREAAQVNSNTSNY 121
QY 121 INAINSKSVSDAINRVSAIREVVSANEXKMLQOQEDKAQVQKQOENQAANTVAANQE 180
Db 122 IDAVLNADSLADAIGRVQAMTTVMKANNDLMEQKQKKAEDKKAENDAKUKELAENQA 181
QY 181 TTAQNTNALNTQQAQLEAQLNLQAEELTTAQQKATLVQAQKAAAEAAQAAAAQAABA 226
Db 182 ALESQKGLLSQAQDLNLVKTSLAAEQATAEKKADLNQKAAEAQAEQARIREQOPLAEQ 241
QY 227 AARQAAAAQAQAAEAKAAAEKALQEAQAQAQAQAANNNTQTADSDA-----DQAAAAAD 279
Db 242 ARQQAQAEKAEKAEQAEAEAE-----QATQASSTAQSSATEESSATQSSMTTESSSATQSS 298
QY 280 NTQAAQGTGSDTQSAQAQAVNNSDQESTTA-----TAAQPSASSASTAAVAANTSS 329
Db 299 ATEESTTPESSTESSTESSTESSTESSTESSTESSTESSTESSTESSTESSTESSTESST 358
QY 330 ANTYPA-----GQCTWGVKSLAP----- 347
Db 359 ESTTPATPTPTDQSDVDGTNGT-GSSTPATPTPTPTPEQPKVTPAPAPSGSVNGAIVAE 417
QY 348 ---WVGN---YWGNG-----GQWASAAAGYRVG-STPSAGA 378
Db 418 AVKYGTPVYVGGKDPGDCSGFTRVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 477
QY 379 VAVW-NDGQYGHVAVTGVQGGQIQVQEAANYAGNQIGNYRGWFGVSVSYIYP 423
Db 478 LLFWGSGQGTTHVATG--GQY-YIHAPQGESVKGVSQV-WFAP 519

RESULT 5
US-09-222-938A-67
; Sequence 67, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match          26.6%; Score 561; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 2e-35;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKKILASLLSTVMVSVQAVLTTHAETTDKIAAQDNKISNLTAQQAQAQVNDIQE 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 61 QVSAIQAEQSNLQAEANDRLQAESKKLEGEITELSKNIVSRNQSLKQARSQAQTNQAVTSY 120

QY 121 INAILNSKVSDAINRVSAIREVVSANERKMLQOQEDKAAVEQKQENQAAINTVAANOE 180
DB 121 INTIIVNSKISITAEISRAVAMSIVSANNKMLQEQKADKKATSEKQVANNDAINTVIANQQ 180

QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTA 210
DB 181 KLADDAQAALTTKQAEKAAELSLAAEKATS 210
```

```
RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990
```

```
Query Match          25.1%; Score 529; DB 2; Length 497;
Best Local Similarity 31.3%; Pred. No. 1.8e-32;
Matches 152; Conservative 83; Mismatches 174; Indels 76; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 16 LKXSVLSALMVCISITLTSVALPSAFADEYDFYKIQOQOKINALTSQMSDAEAKVAAIEN 75

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 76 DMVETAKQIDTTLTKKNKLSSESVSKLYSEISDLNVRQKREVQMTKQARDVQVNGQSDSI 135

QY 121 INAILNSKVSDAINRVSAIREVVSANERKMLQOQEDKAAVEQKQENQAAINTVAANOE 180
DB 136 IDAVIDADSVADAIGRVQAVSTMNANLLEQQEDKATVEKTKNVEKQIAELEAATK 195

QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTA 210
DB 196 ELNDKTESLTKLTKIQEVAKNLDEAQRSEEQKDGFKKXEAELAEQARQRAAK 255

QY 235 -----QAAAEAKAAAEKALQEQAAQAAANNNNTQATDASQQAADNTQAAQTGDS 290
DB 256 KAEQAAAAQAAAAQ- KAAAEQAKATKAA-----NEAASAAAEKAA-----TPVVSSTTT 306
```

```
QY 291 DOSAAQAVNNSDQES-----TTATAAQP-----SASSASTA 321
DB 307 ESTTQETTTSTETESVVTTFVAAAPKEKEVPVNTTPEKNEAKPGNGVTSRQA 366

QY 322 AV-AAANTSSANTYPA-----GOCTWGVKSLAPWVNGYNGGQWMAASAAAAAGYRVG-STP 374
DB 367 AINAAALADVGNYSATGMWQPGCECLVSVRRWLAAAGINFGYGGPNSGYVASGATQVSWSNV 426

QY 375 SAGAVAVNDG-----GYGHVAVVTVGQGOIQVQSEANYAGNQSIGNYRWGFNPGSVS 427
DB 427 QPGDVVQVESAYSPPDSWIGGVHTLVLTGVSGGSVQIVEAN-----NPGSGG 472

QY 428 YIYPN 432
DB 473 YVSSN 477

RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brooks
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 449 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-482

Query Match          23.7%; Score 500.5; DB 2; Length 449;
Best Local Similarity 30.7%; Pred. No. 2.4e-30;
Matches 139; Conservative 88; Mismatches 175; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 VKKRILFASVLLCSLTLSAIAITPSIALADNVDKIEKNQEISLAKAKQDGLASQVSSLEA 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 61 EVSSVFDSESMALREQKTLKAKSEQLQOEITNLNORIEKRNEAIKNQARDVQVNGQSTTM 120

QY 121 INAILNSKVSDAINRVSAIREVVSANERKMLQOQEDKAAVEQKQENQAAINTVAANOE 180
```



```
Query Match      21.4%; Score 452.5; DB 2; Length 422;
Best Local Similarity 30.0%; Pred. No. 1e-26;
Matches 128; Conservative 81; Mismatches 166; Indels 51; Gaps 12;

QY 28 DPFDAIQASQDSKINNLTAAQQAQAVNTIQOVSALOTQQAELQAEQENORLEAQSATLG 87
DB 1 DNVDKKIEEKQEISSLKAKQGDLSQVSSLEAEVSSVFDENVAUREQKOTLKAKSEQLQ 60

QY 88 QOIQTLLSKIVARNESLKQARSQAQSNAAATSVINAIINSKVSVDAINRVSAIREVVSA 147
DB 61 QBITNLNQRIEKNEAIKNQARDVQVNGQSTTMDLDAVLDAQSDVADSAISRQAVSTIVSAN 120

QY 148 EKMLOQEQEDKAAVEQKQEQENQAINTVAANQETIAQNTNALNTQQAQLEAAQLNLQAE 207
DB 121 NDLMQQEKEDQAVVDVKKAEKVKQLEATEAELETKRQDLLSKQSELNVKASLALQ 180

QY 208 TTAQDQKATLVAQKAAAE-----EAAQRAAAQAAAEKAAAEKAKALQEQ 253
DB 181 SSAESSKAGLEKQKAAAEQARLAAEQKAAAEKAKQAAKPAKAEVK--AEAPVASSST 238

QY 254 AQAQAANNNTQATDASDQAAA-----AADNTQAAQTGDSTDQSAQAQAV-----NNSD 302
DB 239 TEAQAPASSS--ATESSQQTETTPSTDNSATENTGSSSEQPVQPTTPSDNGNGG 296

QY 303 QESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNVGGQWMAAS 362
DB 297 QTGGGTVTPTPTPAPSADPTINALNLVRLQSLG-----LRPVV---WDAGLAASAT 345

QY 363 AAAAGYRGVSTPSAGAVAVNNGGCVHVAIVTVGVQGGQIQV---QEAANYAGNQSTGNRYG 419
DB 346 ARAAQVEAGGIPNDH-----WSRG--DEVIAIMWAPGNSVIMAWNETNMVNTASGSG-HRD 398

QY 420 W-FNPG 424
DB 399 WEINPG 404

RESULT 12
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

Query Match      13.4%; Score 283.5; DB 2; Length 469;
Best Local Similarity 28.9%; Pred. No. 1e-13;
Matches 113; Conservative 57; Mismatches 164; Indels 57; Gaps 12;

QY 31 DAQIASQDSKINNLTAAQQAQ-----AAQAVNTIQGVVSALQTOQAE-----LQAEQNR 78
DB 84 DAIMVDPGAVNNYRQQAQQAARSARPAEQKQAQQAQAEELREKQAQAEQERLQLEQER 143

QY 79 LEAQSATLGGQIQTLSSKIVARNESLKQARSQAQSNAAATSVINAIINSKVSVDAINRV 138
DB 144 LQAQEAQAEKEQ-----QKQAEAAAKAAAKAKADAQAKAEQAEAAKAA 190

QY 139 ATRVVVSANEKMLQEQEQDKAAVEQKQEQENQAINTVA--ANQETIAQNTNALNTQQAQLE 197
DB 191 ABAKAKADAQAKAEQAAKAAKAAADAKKQAEAAKAAKAAAEKQAEAAKAAAEKQAKAE 250
```

```
QY 198 AA---QLNLQAEITTTAQDQKATLVAQKAAAEBAARQAAAAQAAAAAEKAAAEKALQEQAA 254
DB 251 AAAAKAQOEAQKQAQEAQKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAA 309

QY 255 QQAQAANNNTQATDASDQQAQAADNTQAAQTGDSTD-----QSAQAQVNNSDQES 305
DB 310 AAQKAAAD---KAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 366

QY 306 TTATAQAQPSASSASTA-----AVAANTSSANTYPAGQCTGWGKSLAP--WVGNV 352
DB 367 AAAAGSGNTKNSAGADINNYAGQIKSAIESKPYDASSYAGKTCTLRK-LAPDGLLLNI 425

QY 353 WNGGQWA-ASAAAAGYRGVSTPSAGAVAV 382
DB 426 QSEGGDPALCQAALAAARQAQKPKPPSQAVY 456

RESULT 13
US-09-134-001C-5035
; Sequence 5035, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5035
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match      11.2%; Score 237; DB 2; Length 264;
Best Local Similarity 34.7%; Pred. No. 1.7e-10;
Matches 60; Conservative 27; Mismatches 70; Indels 16; Gaps 5;

QY 260 ANNNNTQATDASDQQAQAADNTQAAQTGDSTDQSAQAQVNNSDQESTTTATAQPSASSAS 319
DB 104 SNYNNVQ-----SNNTSQRTTQPTGGLGASYSSTSSSNVHVTTTSA-PSSNGVS 151

QY 320 TAAVAANTSSANTYPAGQCTGWV-KSLAPWVGNVWNGGQWAAASAAAAGYRGVSTPSAGA 378
DB 152 LS--NARSASGNLYTSGQCTYYVFDVRVGKIGSTWGNANNANNAARSGYTYNNSPA 209

QY 379 VAVWNDGGYGHVAYVTGV-QGGQIQVOEANYAGNQSIGNYRGWFPNGSVSYIY 430
DB 210 ILQTSQAGYGHVAYVEGVNSNGSIRVSENNYGHGAGVVTSTRTISASQAASVNY 262

RESULT 14
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

Query Match      11.2%; Score 236; DB 2; Length 257;
Best Local Similarity 34.9%; Pred. No. 2e-10;
Matches 60; Conservative 26; Mismatches 70; Indels 16; Gaps 5;

QY 261 NNNNTQATDASDQAAAAADNTQAAQTGDSDDQSAQAQVNNVNSDQESTTATAAOPSSASSAST 320
DB 98 NYNNYQ-----SNNTSQRTTQPTGGLGASYSTSSNVHVTITSA-PSSNGVSL 145

QY 321 AAVAANTSANTYPAGQCTWGV-KSLAPWVGNWNGGOWAAASAAAAGYRVGSTPSAGAV 379
DB 146 S--NARSASGNLYTSGQCTYYVDFRVGGKIGSTWGNANNWANAARSGYTVNNSPAKGAI 203

QY 380 AVWNDGGYGHVAYVTGV-QGGOIQVOEANYAGNQSIGNYRGWFPNPGSVSYIY 430
DB 204 LQTSQAGYGHVAYVEGVNSNGSIRVSEMMYGHGAGVVTSTRTISASQAASINY 255

RESULT 15
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539

Query Match      11.2%; Score 236; DB 2; Length 267;
Best Local Similarity 34.9%; Pred. No. 2.1e-10;
Matches 60; Conservative 26; Mismatches 70; Indels 16; Gaps 5;

QY 261 NNNNTQATDASDQAAAAADNTQAAQTGDSDDQSAQAQVNNVNSDQESTTATAAOPSSASSAST 320
DB 108 NYNNYQ-----SNNTSQRTTQPTGGLGASYSTSSNVHVTITSA-PSSNGVSL 155

QY 321 AAVAANTSANTYPAGQCTWGV-KSLAPWVGNWNGGOWAAASAAAAGYRVGSTPSAGAV 379
DB 156 S--NARSASGNLYTSGQCTYYVDFRVGGKIGSTWGNANNWANAARSGYTVNNSPAKGAI 213

QY 380 AVWNDGGYGHVAYVTGV-QGGOIQVOEANYAGNQSIGNYRGWFPNPGSVSYIY 430
DB 214 LQTSQAGYGHVAYVEGVNSNGSIRVSEMMYGHGAGVVTSTRTISASQAASINY 265

Search completed: February 15, 2006, 18:11:57
Job time : 32.0232 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 92.071 Seconds
(without alignments)
1960.467 Million cell updates/sec

Title: US-10-797-821-31
Perfect score: 2110
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFMFGSVIYPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2110	100.0	432	4	US-10-383-930-31
2	2110	100.0	432	5	US-10-797-821-31
3	2099	99.5	432	5	US-10-383-930-32
4	2099	99.5	432	5	US-10-797-821-32
5	2083.5	98.7	431	4	US-10-383-930-30
6	2083.5	98.7	431	5	US-10-797-821-30
7	2079.5	98.6	431	4	US-10-383-930-29
8	2079.5	98.6	431	5	US-10-797-821-29
9	2075.5	98.4	431	4	US-10-383-930-33
10	2075.5	98.4	431	5	US-10-797-821-33
11	1087	51.5	398	5	US-10-474-792-600
12	1059	50.2	392	5	US-10-472-928-4652
13	1059	50.2	399	5	US-10-617-320-3230
14	588.5	27.9	524	4	US-10-282-122A-57658
15	561	26.6	210	4	US-10-154-251-67
16	500.5	23.7	449	3	US-09-071-035-482
17	500.5	23.7	449	4	US-10-206-576-482
18	500.5	23.7	449	5	US-10-912-362-482
19	452.5	21.4	422	3	US-09-071-035-484
20	452.5	21.4	422	4	US-10-206-576-484
21	452.5	21.4	422	5	US-10-912-362-484
22	325	15.4	630	5	US-10-494-674-6
23	319	15.1	600	3	US-09-738-626-5197
24	302.5	14.3	609	5	US-10-501-282-3184
25	295.5	14.0	440	5	US-10-501-282-3056
26	294	13.9	422	5	US-10-501-282-3054
27	285	13.5	377	5	US-10-501-282-3052

ALIGNMENTS

RESULT 1

US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

Query Match	100.0%;	Score 2110;	DB 4;	Length 432;
Best Local Similarity	100.0%;	Pred. No. 3.1e-119;	Mismatches 0;	Indels 0;
Matches 432;	Conservative 0;	0;	Gaps 0;	0;
QY	1	MKKRILSAVLVSGVTLSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60	
DB	1	MKKRILSAVLVSGVTLSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60	
QY	61	QVSALQTOQAELOAENRLEAQSATLGQIQITLSKIVARNESLKQQAARSQAQSNATSY	120	
DB	61	QVSALQTOQAELOAENRLEAQSATLGQIQITLSKIVARNESLKQQAARSQAQSNATSY	120	
QY	121	INAIINNSVSDAINRVSAIREVVSANEKMLQQQODKAAVSKQKQENQAANTVAANOE	180	
DB	121	INAIINNSVSDAINRVSAIREVVSANEKMLQQQODKAAVSKQKQENQAANTVAANOE	180	
QY	181	TTAONTNALTQQQAQLEAAQLNLQALLETITPAQDKATLVAQKAAAEFAARQAQAQAAEA	240	
DB	181	TTAONTNALTQQQAQLEAAQLNLQALLETITPAQDKATLVAQKAAAEFAARQAQAQAAEA	240	
QY	241	KAAAEKALQEQAQAQAQAANNNNTQATDASQQAQAANTQAAQTGSDTDSQAQAQVNN	300	
DB	241	KAAAEKALQEQAQAQAQAANNNNTQATDASQQAQAANTQAAQTGSDTDSQAQAQVNN	300	
QY	301	SDQESTTATAOPSPASSASTAAVAANTSSANTYPAGOCITGWVKSLAPVGNVYNGNGQWA	360	
DB	301	SDQESTTATAOPSPASSASTAAVAANTSSANTYPAGOCITGWVKSLAPVGNVYNGNGQWA	360	

QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||
QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||
QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||

RESULT 2
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match 100.0%; Score 2110; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 3.1e-119;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
DB ||||||||||||||||||
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
DB ||||||||||||||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB ||||||||||||||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB ||||||||||||||||||
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEQKQENQAANTVAANO 180
DB ||||||||||||||||||
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEQKQENQAANTVAANO 180
DB ||||||||||||||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTA QDOKATLVAQKAAAEARQAAAAQAAAE 240
DB ||||||||||||||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTA QDOKATLVAQKAAAEARQAAAAQAAAE 240
DB ||||||||||||||||||
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB ||||||||||||||||||
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
DB ||||||||||||||||||
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||

RESULT 4
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||
QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||
RESULT 3
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match 99.5%; Score 2099; DB 4; Length 432;
Best Local Similarity 99.3%; Pred. No. 1.4e-118;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
DB ||||||||||||||||||
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
DB ||||||||||||||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB ||||||||||||||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB ||||||||||||||||||
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEQKQENQAANTVAANO 180
DB ||||||||||||||||||
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEQKQENQAANTVAANO 180
DB ||||||||||||||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTA QDOKATLVAQKAAAEARQAAAAQAAAE 240
DB ||||||||||||||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTA QDOKATLVAQKAAAEARQAAAAQAAAE 240
DB ||||||||||||||||||
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB ||||||||||||||||||
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB ||||||||||||||||||
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
DB ||||||||||||||||||
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
DB ||||||||||||||||||
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
DB ||||||||||||||||||
QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||
QY 421 FNPGSVSIYPN 432
DB ||||||||||||||||||

FILE REFERENCE: 25669-020
CURRENT APPLICATION NUMBER: US/10/797,821
CURRENT FILING DATE: 2004-03-09
PRIORITY FILING DATE: 2004-03-09
PRIORITY FILING DATE: 10/383,930
PRIORITY FILING DATE: 2003-03-07
PRIORITY FILING DATE: 60/363,209
PRIORITY FILING DATE: 2002-03-07
PRIORITY FILING DATE: 60/402,483
PRIORITY FILING DATE: 2002-08-08
PRIORITY FILING DATE: 09/290,049
PRIORITY FILING DATE: 1999-04-12
PRIORITY FILING DATE: 60/081,550
PRIORITY FILING DATE: 1998-04-13
PRIORITY FILING DATE: 60/115,142
PRIORITY FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 32
LENGTH: 432
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match 99.5%; Score 2099; DB 5; Length 432;
Best Local Similarity 99.3%; Pred. No. 1.4e-118;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
QY 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAANTVAANQE 180
DB 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAANTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNQLAELTTAQQKATLVAKAAAEAEAAQAAAAA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNQLAELTTAQQKATLVAKAAAEAEAAQAAAAA 240
QY 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTTATAOPSAASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWYNTWNGGQWA 360
DB 301 SQESTTTATAOPSAASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWYNTWNGGQWA 360
QY 361 ASAAAAGYRVGSTPSAGAVAVWVDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVWVDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPQSVSYIYPN 432
DB 421 FNPQSVSYIYPN 432

RESULT 5
US-10-383-930-30
Sequence 30, Application US/10383930
Publication No. US20040127400A1
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J
APPLICANT: Taubman, Martin A
TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
FILE REFERENCE: 25669-018
CURRENT APPLICATION NUMBER: US/10/383,930
CURRENT FILING DATE: 2003-03-07
PRIORITY FILING DATE: 60/402,483
PRIORITY FILING DATE: 2002-08-08

PRIOR APPLICATION NUMBER: 60/363,209
PRIORITY FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.2
SEQ ID NO 30
LENGTH: 431
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-383-930-30

Query Match 98.7%; Score 2083.5; DB 4; Length 431;
Best Local Similarity 99.1%; Pred. No. 1.2e-117;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
QY 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAANTVAANQE 180
DB 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAANTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNQLAELTTAQQKATLVAKAAAEAEAAQAAAAA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNQLAELTTAQQKATLVAKAAAEAEAAQAAAAA 240
QY 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTTATAOPSAASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWYNTWNGGQWA 360
DB 301 SQESTTTATAOPSAASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWYNTWNGGQWA 360
QY 361 ASAAAAGYRVGSTPSAGAVAVWVDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVWVDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPQSVSYIYPN 432
DB 421 FNPQSVSYIYPN 431

RESULT 6
US-10-797-821-30
Sequence 30, Application US/10797821
Publication No. US20050031633A1
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J
APPLICANT: Taubman, Martin A
TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
FILE REFERENCE: 25669-020
CURRENT APPLICATION NUMBER: US/10/797,821
CURRENT FILING DATE: 2004-03-09
PRIORITY FILING DATE: 10/383,930
PRIORITY FILING DATE: 2003-03-07
PRIORITY FILING DATE: 60/363,209
PRIORITY FILING DATE: 2002-03-07
PRIORITY FILING DATE: 60/402,483
PRIORITY FILING DATE: 2002-08-08
PRIORITY FILING DATE: 09/290,049
PRIORITY FILING DATE: 1999-04-12
PRIORITY FILING DATE: 60/081,550
PRIORITY FILING DATE: 1998-04-13
PRIORITY FILING DATE: 60/115,142
PRIORITY FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 30

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match          98.7%; Score 2083.5; DB 5; Length 431;
Best Local Similarity 99.1%; Pred. No. 2.1e-117;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAQKAAAEBAARQAAAAQA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAQKAAAEBAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTSDSAAQAVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTSDSAAQAVNN 300
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTSDSAAQAVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTSDSAAQAVNN 300
QY 301 SDOESTTATAOPSASSASTAAVAANTSANTYPAGQCTGWKSLAPWVGNWNGGQWA 360
Db 301 SDOESTTATAOPSASSASTAAVAANTSANTYPAGQCTGWKSLAPWVGNWNGGQWA 360
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
Db 421 FNPGSVSYIYPN 431

RESULT 8
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match          98.6%; Score 2079.5; DB 5; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.1e-117;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180

; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match          98.6%; Score 2079.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.1e-117;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180

; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match          98.6%; Score 2079.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.1e-117;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAATNTVAANOE 180
```

Db 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQOENQAINNTVAANQE 180
QY 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
QY 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 300
Db 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 299
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360
Db 300 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419
QY 421 FNPGSVSYIYPN 432
Db 420 FNPGSVSYIYPN 431
RESULT 9
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33
Query Match 98.4%; Score 2075.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 3.7e-117;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSSTATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQOENQAINNTVAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQOENQAINNTVAANQE 180
QY 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
QY 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 300
Db 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 299
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360
Db 300 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419

Db 300 SDOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419
QY 421 FNPGSVSYIYPN 432
Db 420 FNPGSVSYIYPN 431
RESULT 10
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33
Query Match 98.4%; Score 2075.5; DB 5; Length 431;
Best Local Similarity 98.6%; Pred. No. 3.7e-117;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSSTATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQOENQAINNTVAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQOENQAINNTVAANQE 180
QY 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAKKAAAEBAARQAAAAQAAAE 240
QY 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 300
Db 241 KAAAEAKALQEQAQAQAQAANNNTQTATDASDQAAAAADNTQAAQTGDSFGSAQAQVNN 299
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360
Db 300 SDOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419


```

; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/617,320
;   FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
;   APPLICATION NUMBER: 60/ 085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneka
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 399 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (8) LOCATION 1...399
;     SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230

Query Match      50.2%; Score 1059; DB 5; Length 399;
Best Local Similarity 50.0%; Pred. No. 5.9e-56;
Matches 218; Conservative 75; Mismatches 91; Indels 52; Gaps 4;

QY      1  MKKRILSAVLVSGVTLSATTLSAISKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db      8  MKKILASLLSTVMVSQVAVLTTHAHTTDDKTAADNKISNLTAAQQEAQKQVDIQE 67
QY      61 QVSALQTOQAELOAENQRLAQSATLGGQIOTLTSSKIVARNESLKQARSQAQKSNAAATSY 120
Db      68 QVSAIQAEQSNLQANDRLQAESKKLEIGETLSKNIIVSRNQSLEKQARSQAQTNQAVTSY 127
QY      121 INAINSKSVSDAINRVSAIREVVSANEKMLQQEQDQAAVEQKQEQENQAINVVAANOE 180
Db      128 INTIVNSKITEAISRVAAVMSSEIVSANNKMLEQQKKKAISEKQVANNDAINTVIANQQ 187
QY      181 TTAQNTNALNTQOAELOAENQRLAQSATLGGQIOTLTSSKIVARNESLKQARSQAQKSNAAATSY 240
Db      188 KLADAQALTTQAEKKAELSLEAKETATGEKASLLEQLAAAEAEARAAVAEAAAYKE 247
QY      241 KAAAEAKALQEAQAQAQAANNNTQATDASDQQAQAADNTQAAQTGSDTQSAQAQVNN 300
Db      248 KRASQQQSV-----LASANTLTAAQVQAVSAGAAAPVRKVRPT----- 286
QY      301 SQDESTTATAQPSASSASTRAVAANTSSANTYPAGQCTGWGKSLAPWGVNKGNGQWA 360
Db      287 -----YSTNSSYPIGECTGWGKTLAPWAGDYGNGAQWA 321
QY      361 ASAAAAGYRVGTPSAGAVAVNDDGGYGHVAVVTGVQG-GQIQVQEAANYAGNOSIGNYRG 419
Db      322 TGAAGAAGFTGTPQVGAIAACWNDGGYGHVAVVTAVESTTRTIQVSESNYAGNRTIGNHRG 381
QY      420 WFNP-----GSVSYIY 430
Db      382 WFNPTTSEGFTYIY 397

RESULT 14
US-10-282-122A-57658

```

RESULT 14
US-10-282

Db	298	ATEESTTPESSTESSTAPESSTATEESTTAPESSTATEESTTVPESSTATEESTTVPESSTTE	357
Qy	330	ANTVPA-----GQCTGWGKSLAP-----	347
Db	358	ESTTPAPTPTDQSDVDGNGT--GSSTPAPTPTPEQPKVTPAPAPSGSVNGAAIVAE	416
Qy	348	---WVGN--YVNGN-----GWAASAAAAGYRVG--STPSAGA	378
Db	417	AYKIGTPTVYWGKDPGDCSGFTRYVYMQVTGRDIGGWTVPQESAGTKVISOAKAGD	476
Qy	379	VAVW-NDGGYGHVAVVTGVGGQIQVQAEANYAGNQSIGNYRQWFPN	423
Db	477	LLFWGSOGGTHVAIALG--GGQ-YIHAPQGESVKVGSVQ--WFAF	518

RESULT 15

```

US-10-154-251-67
; Sequence 67, Application US/10154251
; Publication No. US20030092024A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL
; FILE REFERENCE: 06286-060002
; CURRENT APPLICATION NUMBER: US/10/154
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 10/154
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-154-251-67

```

```
Query Match      26.6%; Score 561; DB 4; Length 210;
Best Local Similarity 55.2%; Pred. No. 2.8e-26;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;
```

Qy	1	MKKRILSAVLVGVVTLTSATTLSAIKADFDRAQIASQDSKINNLTAAQQAQAQVNTTIG	60
Db	1	MKKKILASLLSTVMVSVQAVLTTTAHETDDDKIAAQDNKISLNTAAQQAQKQVDIOE	60
Qy	61	QVSALOTQQAELQAEQNRLEAQSATLGOOIOTLSSKIVARNESLQQAARSAQSNAAYS	120
Db	61	QVSAIQAEQSNLQAEQNRLEAQSCKLKEGIELTSKNI VSRNQSLKQAASQATNGAYSY	120
Qy	121	INATINSKVSADAINRVSAIREVVSANESKMLQQOQEQDKAAAEVQKQOENQAALINTVAANOE	180
Db	121	INTIVNSKSITEAISRVAAEMSEIVSANNKMLEQKQADKKAKISEKVANNDAINTVIANQQ	180
Qy	181	TIQNTNALNTQQQAQLEAAQLNLQAEILTTA	210
Db	181	KLADDAQALATTKQAELKAAELSLAAEKATS	210

Search completed: February 15, 2006, 18:34:52
Job time : 94.071 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:30:27 ; Search time 8.33977 Seconds
(without alignments)
736.166 Million cell updates/sec

Title: US-10-797-821-31
Perfect score: 2110
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGMFNGSVIYIPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubpaa/US05_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079.5	98.6	431	7	US-11-052-554A-210 Sequence 210, App
2	1093	51.8	398	7	US-11-052-554A-252 Sequence 252, App
3	436.5	20.7	211	7	US-11-052-554A-352 Sequence 352, App
4	411	19.5	544	7	US-11-052-554A-358 Sequence 358, App
5	236	11.2	257	6	US-10-793-626-3244 Sequence 3244, App
6	229.5	10.9	394	7	US-11-052-554A-79 Sequence 79, Appl
7	229	10.9	971	7	US-11-052-554A-3 Sequence 3, Appl1
8	226	10.7	1236	6	US-10-873-528-109 Sequence 109, App
9	222.5	10.5	270	6	US-10-485-517-413 Sequence 413, App
10	214	10.1	149	6	US-10-793-626-1692 Sequence 1692, App
11	214	10.1	157	6	US-10-793-626-2870 Sequence 2870, App
12	191.5	9.1	1562	7	US-11-052-554A-211 Sequence 211, App
13	191.5	9.1	2101	6	US-10-857-780-23 Sequence 23, Appl
14	187.5	8.9	1095	6	US-10-793-626-3154 Sequence 3154, App
15	186.5	8.8	1586	6	US-10-821-234-901 Sequence 901, App
16	186	8.8	1448	6	US-10-485-517-212 Sequence 212, App
17	184.5	8.7	758	6	US-10-485-517-144 Sequence 144, App
18	184.5	8.7	1410	6	US-10-878-556A-136 Sequence 136, App
19	181.5	8.6	1122	6	US-10-467-657-6112 Sequence 6112, App
20	181	8.6	330	6	US-10-485-517-413 Sequence 413, App
21	178.5	8.5	1565	6	US-10-467-657-2704 Sequence 2704, App
22	178	8.4	1107	6	US-10-485-517-145 Sequence 145, App
23	178	8.4	1290	6	US-10-485-517-141 Sequence 141, App
24	175	8.3	1126	6	US-10-485-517-248 Sequence 248, App
25	174	8.2	834	7	US-11-052-554A-212 Sequence 212, App

ALIGNMENTS

RESULT 1

US-11-052-554A-210
; Sequence 210, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589, 227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 210
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-210

Query Match 98.6%; Score 2079.5; DB 7; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.3e-112;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy	1	MKKRILSAVLVSGVTLSATLSAKDDFDAIASQDSKINNLTAAQQAAQAQVNTIOG	60
Db	1	MKKRILSAVLVSGVTLSATLSAKDDFDAIASQDSKINNLTAAQQAAQAQVNTIOG	60
Qy	61	QVSALQTQAELEAQNRLAQSAATLGGQIQTLSSKIVARNESLKQQAASAKSNAATSY	120
Db	61	QVSALQTQAELEAQNRLAQSAATLGGQIQTLSSKIVARNESLKQQAASAKSNAATSY	120
Qy	121	INAINSKSVDAINRVSAIREVVSANEKMLQQQODKAAVQKQENQAANTVAANOE	180
Db	121	INAINSKSVDAINRVSAIREVVSANEKMLQQQODKAAVQKQENQAANTVAANOE	180
Qy	181	TTAONTNALNTQAAQLEAAQLNLQAELETTAODKATLVAOKAAAEAAQAAQAQAAEA	240
Db	181	TTAONTNALNTQAAQLEAAQLNLQAELETTAODKATLVAOKAAAEAAQAAQAQAAEA	240
Qy	241	KAAAEAKALQEAQAQAQAAANNNTQATDASDQQAADNTQAQTGSDTDSQAQAVNN	300
Db	241	KAAAEAKALQEAQAQAQAAANNNTQATDVSDDQAADNTQAQTGSDTDSQAQAVNN	299
Qy	301	SDQESTTATAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWGVNTWNGGQWA	360

Db 300 SDOESTTATEAPSPASSASTRAVAANTSSANTYPAGQCTGWKSLAPWGVNMGQWA 359
QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGGQIQVQBEANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGGQIQVQBEANYAGNOSIGNYRGW 419
QY 421 FNPGSVSYIYPN 432
Db 420 FNPGSVSYIYPN 431

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match 51.8%; Score 1093; DB 7; Length 398;
Best Local Similarity 53.8%; Pred. No. 5e-56;
Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;
QY 1 MKKRILSAVLVSGVTLSSTATTLSAIKADDFDQAIASQSKINNLTAAQQAQAQVNTTQ 60
Db 1 MKKRILSAVLVSGVTILGAATT---VGAEDLSKIAKQDSIISNLATTEQKAAQNOVSALQA 57
QY 61 QVSALQTOQAELOAENORLEAQSATILGOQIOTLSSKIVARNESLKQOARSACKSNAATSY 120
Db 58 QVSSIQSEODKUTARNTLEALSKEPQEIKAITSQI VARNELKLNQARSATYKNETSGY 117
QY 121 INAIINSKVSDAINRVSAIREVWSANERKMLQQQEQDKAAVEQKQEQENQAANTVVAANQE 180
Db 118 INALLNSKISIDVNLVAINRAVSANAKLLSQKADKVSLEKQAAQNTAINTIAANWA 177
QY 181 TTAQNTNALNTQOALEAAQLNLQELTAAQDKATLVQAQKAAABEAAQAAAAQAAAAEA 240
Db 178 MAEENQNTLRTQANLEATANTLALQLASATEDKANLVAQKAAEAKAAEAALAQQAARV 237
QY 241 KAAAEKALOEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQDQSAQAQVNN 300
Db 238 KA-----QEQA-----QQAASVEAKSAITPAPQATPAAQSSNA 272
QY 301 SDOESTTATAAPSPASSASTAAVAANTSSANTYPAGQCTGWKSLAPWGVNMGQWA 360
Db 273 LEPAALTAPAA-PSARPQT-----SYDSSNTYPVPGQCTGWAKSLAPWAGNMGQWA 325
QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGQIQVQBEANYAGNOSTGNVTRG 419
Db 326 YSAQAAGYRTGTPMVGAIVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYIADHRG 385
QY 420 FNPGSVSYIYPN 432
Db 386 WFNPVTGVTIYPH 398

RESULT 3

US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match 20.7%; Score 436.5; DB 7; Length 211;
Best Local Similarity 44.2%; Pred. No. 7.9e-19;
Matches 95; Conservative 25; Mismatches 60; Indels 35; Gaps 7;
QY 247 KALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDS-----TDQSAQA 297
Db 3 KQLEKAVFTVAA-----TAATVVLGNKWDAD-TYTLQEGDGFPSVAQRVHMDAYELAS 56
QY 298 VNNSD-----QBSTTATAAQAQPSASSA-----STAAVAANTSSANTYPAGQCTGWKVS 344
Db 57 MNGKDIITSLILPGQTLTVNGSAAPDQAAAAPTDTQTATTENDANANTYPVQCTGWKVA 116
QY 345 LAPWGVNMGVNGGOWAASAAAGYRVGTPSAGAVVNDGGYGHVAVTVGV-QGSGIQV 403
Db 117 VATWAGDWGNGGOWASSASAAQGYTVGNTPAVGSIMCMTDGGYGHVAVTVAVGEDGKVQV 176
QY 404 QEANYAGNOSIGNYRGWEN-----PGSVSYIYPN 432
Db 177 LESNYKDOQWVDNVRGWDFDPNNSGTPGSVSYIYPN 211

RESULT 4
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match 19.5%; Score 411; DB 7; Length 544;
Best Local Similarity 25.3%; Pred. No. 6.7e-17;
Matches 143; Conservative 78; Mismatches 186; Indels 158; Gaps 19;
QY 4 RILSAVLVSGVTLSSATLSAIAKDDFPAQTA-----SQDSKINNLTAAQQAQAQVNTIQ 59
Db 2 RKLKVALFASSILGMLAVSSYTAADTEDNQVTISHYNEQAGTFDYNVQAANGKTIQSID 61


```
Db 187 STYWSDAKWASNAANDGYQVNDTPSVGAIMOSTPGPYGHVAYVERINGDSILISEMNY 246
QY 409 AGNQSIGNRY 418
Db 247 ANGPYNMYR 256

RESULT 10
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match 10.1%; Score 214; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2.8e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSSTDSAAQAVNNSDOESTTATAAOPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345
Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDYKT 67

QY 346 APWVGNYMGNGGOWAASAAAGRYVGSTPSAGAVAVMNDGGYGHVAYVTGVQG-GOIQVQ 404
Db 68 GKGISTWGNANSWATAAQAAGFTVNTTPEGAIMQSSEGAFGHVAFVSVNNDGSITVS 127

QY 405 EANYAG 410
Db 128 EMNYDG 133

RESULT 11
US-10-793-626-2870
; Sequence 2870, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2870
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2870

Query Match 10.1%; Score 214; DB 6; Length 157;
Best Local Similarity 39.7%; Pred. No. 2.9e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;
```

```
QY 287 GDSSTDSAAQAVNNSDOESTTATAAOPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345
Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDYKT 67

QY 346 APWVGNYMGNGGOWAASAAAGRYVGSTPSAGAVAVMNDGGYGHVAYVTGVQG-GOIQVQ 404
Db 68 GKGISTWGNANSWATAAQAAGFTVNTTPEGAIMQSSEGAFGHVAFVSVNNDGSITVS 127

QY 405 EANYAG 410
Db 128 EMNYDG 133

RESULT 12
US-11-052-554A-211
; Sequence 211, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 211
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-211

Query Match 9.1%; Score 191.5; DB 7; Length 1562;
Best Local Similarity 26.3%; Pred. No. 0.00078;
Matches 87; Conservative 58; Mismatches 115; Indels 71; Gaps 13;

QY 28 DDFDAQIASQDSKINNLTAAQQAAQAQVNTIQGVSAQTQQAELQAEQNRLEAQSAT-- 85
Db 151 DQYKSDVAHAHEAVEAVAKIRAKNQATKEQ-----YKDMAAHKAEVERINAANAASK 200

QY 86 -----LGOQIQTLSS--KIVARNESLKKQARSA-----QKSNAA--TSYINAI1-- 125
Db 201 TAYEAKLAQYQADLAQVKTNAQAAQYQKALAAQYQAEKLVQENAAKAAAYDTAVAN 260

QY 126 NSKSVSDAINRVSAIREVVVSANEKMLQQOQEQDKAAVEQKQENQAAINTVAANQETIAQN 185
Db 261 NAKNT-----ETAAANEETIRKNATAKAEYETKLAQYQAEKRV-----QE 301

QY 186 TVALNTQQAQLAAQLNLQAEILTAAQDQKATLVQAKAAAEAAARQAAQAAAEAKAAAE 245
Db 302 ANAAN-----EADYQAKLTAYQTELARVOKANADAKAAAYEAFAVANNNAALTAETAE 352

QY 246 AKALQEQAAQAAA--ANNNTQATDASDQAAAAADNT--QAAQTGDSSTDQSAQAQVNNNS 301
Db 353 NTAIKQRNENAKATYEALKKQYEDLAALAVKKANANEADYQAKLTAYQTELARVOKAN-- 410

QY 302 DQESTTATAAOPSASSASTAAVAANTSSANT 332
Db 411 ----ADAKAAAYEAFAVANNNAALTAETAE-NT 436

RESULT 13
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
```

```

; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEROEF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR FILING DATE: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-23

Query Match          9.1%; Score 191.5; DB 6; Length 2101;
Best Local Similarity 22.0%; Pred. No. 0.0011;
Matches 90; Conservative 68; Mismatches 141; Indels 111; Gaps 15;

QY 2 KKRILSAVLGVLTSSATTLISAIRAD-----DFDAIASODSKINNLTAAQQA 51
Db 460 EQQLSSLTID--LOSSISNLQAKKELEQASQAHGARTLAQVASLTSELTTLNA---TI 514
QY 52 QAVNTTIOGVVSALTOQAELQAEORLEAQSATLGGQIQTLSKIVARNESLKKOARS- 110
Db 515 OQDQDELGLKQAKKEKQAQLAQTTLQOQEQASQGLRHQVEQLSSLKQKEQQLKEVAEKQ 574
QY 111 -----AQKSNAASTYNAIINSKVSVDINRVSAREVWSANEKMLQQ----- 154
Db 575 EATRDHQAQLATAAEEERASLRERDA--ALKQLEALEKEKAALKLEILQOQLQVANEARD 632
QY 155 -----EQDKAAVEQKQBNQAAINTVAANQETIAQNTNALNTQQAQLPAAQLNLQA 205
Db 633 SAQTSVTOQAQKAEKSLRKVEELQACVET--ARQE-----QHEAQAQVASELQL 680
QY 206 ELTTAQDQKAT-----LVAQKAAAEAEARQAAQAAAEAKAA 243
Db 681 R---SEQQKATEKERVEAQEKDQLQELQALQKESLKVTKGSLEEEKRRADA-----LE 730
QY 244 AEAKALQEAQAQAQAANNNNTQATDASDQQAAD--NTQAAQTGSDTSDQSAQAQVNSD 302
Db 731 EQQRICISLKAETSLVQHKREKELEERAGRKGLERLLQLGEA-HQATEVLR--- 786
QY 303 QESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGTVKSLAPVNGY 352
Db 787 RELAEAMAQAHTAES-----ECEQLVKEVAAWRDGY 817

RESULT 14
US-10-793-626-3154
; Sequence 3154, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3154
; LENGTH: 1095
; TYPE: PRT

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3154

Query Match          8.9%; Score 187.5; DB 6; Length 1095;
Best Local Similarity 22.7%; Pred. No. 0.0088;
Matches 93; Conservative 59; Mismatches 166; Indels 91; Gaps 13;

QY 13 GVTLSATTLISAIRADDFDAIASODSKINNLTAAQQAQAVNTTIOGVVSALOTQAEL 72
Db 340 GWTDTTANNYKSKKEAED-ELQKAQIINNNGDAFEQQITNETNRVNOAINAKAKNDL 398
QY 73 QAEORLEAQSATLGGQIQTLSKIVA-RNESLKKOARSQAQKSNAASTYNAIINS---- 127
Db 399 RADKSQLENAYNQLTQNVDTNGKKPASTIQYQAARQAIETQYNNAKSEAHQILENSPSV 458
QY 128 KVSVDINRVSAREVWSANEKMLQQQEQDQKAAVSEKQEQENQA-----AINT 174
Db 459 NEVAQALQKVEAVQLKVNDAITHMLQKNNKNSALVTAKLQLOQAVNDQPLTTGMTODSINN 518
QY 175 VAA-----NOETIAQNTNA-----LNTQQAQLEAAQLNLQ-----BELTTA 210
Db 519 YVAKRNEAQSARNAEAVINNGDATAKQISDEKSKVEQALAHNDKAKQOLTADTTTELQTA 578
QY 211 QDQ-----KATLVAQKAAAEAEARQAAQAAAEAKAAAEAKALQEQAAQANA 259
Db 579 VOOLNRGDTNNKKPRNSINAYNKAIQSLETQITSADKNANAVIQPIRTVQE-VNNALQQ 637
QY 260 ANNNNTQATDASDQQAADN-----TQAAQTGSDTSDQ----- 293
Db 638 VNQLNQQLTEALINQLPLSNNDALKAAARLNLENKINQIVQTDGTTQQSIEAYQNAKRVQA 697
QY 294 -----AAQAVNSD---QESTTATAQPSASSASTAAVAANTSSANTYP 334
Db 698 NESNTALALINNGDAEQIITETDRVNQQTTLNLTOAI--NGLTVNKEP 744

RESULT 15
US-10-821-234-901
; Sequence 901, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 901
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-901

Query Match          8.8%; Score 186.5; DB 6; Length 1586;
Best Local Similarity 22.3%; Pred. No. 0.0015;
Matches 86; Conservative 77; Mismatches 124; Indels 99; Gaps 16;

QY 33 QIASODSKINNLTAAQQAQ---AQVNTTIOGVVSALQ-----TQAEIQAENQRL- 79
Db 923 QLVAREQETAVQAEMQASVREHVKEVQQLQKIRTLQEQLENGPNTQLARLQENSIILR 982
QY 80 -----EAQSATLGGQIQTLSKIVARNESLQ---QARSQKSNAASTYNA 123
Db 983 DALNQATSOVESKQNAELAKRLQELSKVSKELVEKSEAVRQDEQQQKALEAKAA-AFEKQ 1041

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.754 Seconds
(without alignments)
1668.609 Million cell updates/sec

Title: US-10-797-821-32
Perfect score: 2116
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGWNPGSVSIYIPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2116	100.0	432	7	ADD93652 Streptoco
2	2116	100.0	432	9	ADx37275 Streptoco
3	2099	99.2	432	7	ADD93651 Streptoco
4	2099	99.2	432	9	ADx37274 Streptoco
5	2074.5	98.0	431	7	ADD93650 Streptoco
6	2074.5	98.0	431	9	ADx37273 Streptoco
7	2070.5	97.8	431	9	ADx37272 Streptoco
8	2070.5	97.8	431	9	ADx37272 Streptoco
9	2070.5	97.8	431	9	ADx37272 Streptoco
10	2066.5	97.7	431	7	ADD93653 Streptoco
11	2066.5	97.7	431	9	ADx37276 Streptoco
12	1229.5	58.1	447	5	ADx37276 Streptoco
13	1229.5	58.1	447	8	ADU69524 S agalact
14	1229.5	58.1	447	8	ADU69524 S agalact
15	1229.5	58.1	447	8	ADU69524 S agalact
16	1229.5	58.1	447	8	ADU69524 S agalact
17	1087	51.4	398	9	ADx37276 Streptoco
18	1081	51.1	398	5	ADx37276 Streptoco
19	1081	51.1	398	8	ADx37276 Streptoco
20	1071.5	50.6	395	5	ADx37276 Streptoco
21	1053	49.8	392	6	ADx37276 Streptoco
22	1053	49.8	392	8	ADx37276 Streptoco
23	1053	49.8	392	8	ADx37276 Streptoco
24	1053	49.8	399	8	ADx37276 Streptoco

25	1053	49.8	399	9	AEA58465	Aea58465 Streptoco
26	1052	49.7	392	8	ADK47859	Adk47859 Streptoco
27	1052	49.7	392	8	ADT50227	Adt50227 S_pneumon
28	679.5	32.1	461	2	AAR14150	Aar14150 MSP encod
29	678.5	32.1	461	2	AAR14530	Aar14530 Usp45 pro
30	658	31.1	456	5	ABB55584	Abb55584 Lactococc
31	585.5	27.7	524	6	ABU29734	Abu29734 Protein e
32	581.5	27.5	525	7	ADC95468	Adc95468 E. faeciu
33	555	26.2	210	2	ADY22579	Ady22579 Bacterial
34	527	24.9	482	9	ADV16553	Adv16553 E. faecal
35	527	24.9	497	7	ADH88105	Adh88105 Enterococ
36	494.5	23.4	449	5	ABP43469	Abp43469 E. faecal
37	494.5	23.4	449	5	ABP43469	Abp43469 E. faecal
38	494.5	23.4	449	6	ABU88497	Abu88497 E. faecal
39	494.5	23.4	449	6	ABU13748	Abu13748 Enterococ
40	494.5	23.4	449	9	ADV16734	Adv16734 E. faecal
41	494.5	23.4	449	9	ADY39216	Ady39216 Novel Ent
42	494.5	23.4	450	7	ADH87829	Adh87829 Enterococ
43	445.5	21.1	422	2	AAU00251	Aay00251 Enterococ
44	445.5	21.1	422	5	ABP43470	Abp43470 E faecali
45	445.5	21.1	422	6	ABU88498	Abu88498 E. faecal

ALIGNMENTS

RESULT 1
ADD93652
ID ADD93652 standard; protein; 432 AA.
XX
AC ADD93652;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
(FORS-) FORSYTH INST.
Smith DU, Taubman MA;
WPI: 2003-845091/78.
GENBANK; AY046413.

Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
Claim 5; Page 8; 49pp; English.
The present sequence is the protein sequence of the glucan binding protein-B (GbpB) of Streptococcus mutans strain 3S1. The sequence includes a number of human leucocyte antigen (HLA)-binding epitopes. The invention provides immunogenic compositions and vaccines for dental caries. The compositions comprise major histocompatibility complex (MHC) class II protein-binding GbpB peptides, especially HLA-binding peptides, covalently linked with peptide subunits of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiptopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of Gbps can be used in

XX 07-MAR-2003; 2003WO-US006962.
XX PF
XX 07-MAR-2002; 2002US-0363209P.
XX PR
XX 08-AUG-2002; 2002US-0402483P.
XX PR
XX (FORS-) FORSYTH INST.
XX PA
XX Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
XX GENBANK; AY046412.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 5; Page 8; 49pp; English.
XX
XX The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 15Jp2. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Dieptopic or multiepitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
XX Sequence 432 AA;

Query Match 99.2%; Score 2099; DB 7; Length 432;
Best Local Similarity 99.3%; Pred. No. 1.1e-133;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQQAARSQAQKSNAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQQAARSQAQKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQDQKAAVEQKHQENQAATNTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240
QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAAAQAVNN 300
DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAAAQAVNN 300
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGCCTWGYKSLAPWVNTWNGGOWA 360
DB 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGCCTWGYKSLAPWVNTWNGGOWA 360
QY 361 ASAAAAGRVGSTPAGAVVWVNDGGYGHVAVVTGVQGGQIQVQENYAGNQSIGNYRGW 420
DB 361 ASAAAAGRVGSTPAGAVVWVNDGGYGHVAVVTGVQGGQIQVQENYAGNQSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 432

ID ADX37274 standard; protein; 432 AA.
XX
AC ADX37274;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #3.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-0029004P.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 3; SEQ ID NO 31; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 432 AA;

Query Match 99.2%; Score 2099; DB 9; Length 432;
Best Local Similarity 99.3%; Pred. No. 1.1e-133;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQQAARSQAQKSNAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQQAARSQAQKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQDQKAAVEQKHQENQAATNTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240
QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAAAQAVNN 300
DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAAAQAVNN 300

QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPVWNTWNGGOWA 360
 DB 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPVWNTWNGGOWA 360
 QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEBYVAGNOSIGNYRGW 420
 DB 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEBYVAGNOSIGNYRGW 420
 QY 421 FNPGSVSYIYPN 432
 DB 421 FNPGSVSYIYPN 432

RESULT 5
 ADD93650
 ID ADD93650 standard; protein; 431 AA.
 AC ADD93650;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Streptococcus mutans glucan binding protein-B.
 XX
 KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
 XX
 OS Streptococcus mutans.
 XX
 PN WO2003075845-A2.
 PD 18-SEP-2003.
 PF 07-MAR-2003; 2003WO-US006962.
 XX
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 XX
 PA (FORS-) FORSYTH INST.
 XX
 PI Smith DJ, Taubman MA;
 FI WPI; 2003-845091/78.
 DR GENEBAK; AY046411.
 XX
 PT Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.
 XX
 PS Claim 5; Page 8; 49pp; English.
 XX
 CC The present sequence is the protein sequence of the glucan binding
 CC protein-B (GbpB) of Streptococcus mutans strain 3VP4. The sequence
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
 CC invention provides immunogenic compositions and vaccines for dental
 CC caries. The compositions comprise major histocompatibility complex (MHC)
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
 CC covalently linked with peptide subunits of a glucosyltransferase. The
 CC compositions are used in a claimed method of eliciting production of an
 CC antibody in a mammal. Diapetopic or multi-peptide polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.
 XX
 SQ Sequence 431 AA;

Query Match 98.0%; Score 2074.5; DB 7; Length 431;
 Best Local Similarity 98.8%; Pred. No. 5.1e-132;
 Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
 DB 1 MKKRILSAVLVSGVTLSSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
 QY 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

DB 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
 QY 121 INAIINSKVSDDAINRVSAIREVVSANEKMLHQEQBDKAAVEQKHQENQAINTVAANQE 180
 DB 121 INAIINSKVSDDAINRVSAIREVVSANEKMLHQEQBDKAAVEQKHQENQAINTVAANQE 180
 QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAAARQAAAAQAAEA 240
 DB 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAAARQAAAAQAAEA 240
 QY 241 KAAAEKALQEQAAQAAQAAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300
 DB 241 KAAAEKALQEQAAQAAQAAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300
 QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPVWNTWNGGOWA 360
 DB 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPVWNTWNGGOWA 360
 QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEBYVAGNOSIGNYRGW 420
 DB 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEBYVAGNOSIGNYRGW 420
 QY 421 FNPGSVSYIYPN 432
 DB 421 FNPGSVSYIYPN 432

RESULT 6
 ADX37273
 ID ADX37273 standard; protein; 431 AA.
 AC ADX37273;
 XX
 DT 21-APR-2005 (first entry)
 XX
 DE Streptococcus mutant glucan binding protein B variant #2.
 XX
 KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX
 OS Streptococcus mutans.
 XX
 PN US2005031633-A1.
 XX
 PD 10-FEB-2005.
 XX
 PF 09-MAR-2004; 2004US-00797821.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 98US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 XX
 XX (SMIT/) SMITH D J.
 PA (TAUB/) TAUBMAN M A.
 XX
 PI Smith DJ, Taubman MA;
 XX
 DR WPI; 2005-151644/16.
 XX
 PT New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.
 XX
 PS Claim 3; SEQ ID NO 30; 73pp; English.
 XX
 CC The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II

CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 431 AA;

Query Match 98.0%; Score 2074.5; DB 9; Length 431;
Best Local Similarity 98.8%; Pred. No. 5.1e-132;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQITLSSKIVARNESLKQQAARSQAQKSNAATSY 120
DB 61 QVSALQTOQAELQAEORLEAQSATLGGQIQITLSSKIVARNESLKQQAARSQAQKSNAATSY 120

QY 121 INAINSKSVSDAINRVSAREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180
DB 121 INAINSKSVSDAINRVSAREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180

QY 181 TTAQNTNALNTQQAQLLEAQLNLQAEITTAQDQKATLVAKKAAEEAARQAAAAQAAAAE 240
DB 181 TTAQNTNALNTQQAQLLEAQLNLQAEITTAQDQKATLVAKKAAEEAARQAAAAQAAAAE 240

QY 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300

QY 301 SQEESTATAQPSASASASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360
DB 301 SQEESTATAQPSASASASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360

QY 361 ASAAAGYRGVSPSAGAVVNDGGYGVYVTVGGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAGYRGVSPSAGAVVNDGGYGVYVTVGGGQIQVQEAANYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 7
ADD93649
ID ADD93649 standard; protein; 431 AA.
XX
AC ADD93649;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
FH Key Location/Qualifiers
FT Region 6..25 /note= "HLA-binding peptide"
FT Region 16..35 /note= "HLA-binding peptide"
FT Region 33..52 /note= "HLA-binding peptide"
FT Region 37..56 /note= "HLA-binding peptide"
FT Region 48..67 /note= "HLA-binding peptide"
FT Region 52..71 /note= "HLA-binding peptide"
FT Region 88..107 /note= "HLA-binding peptide"
FT Region 113..132 /note= "HLA-binding peptide"

FT Region /note= "HLA-binding peptide"
FT 117..136 /note= "HLA-binding peptide"
FT Region 137..156 /note= "HLA-binding peptide"
FT 174..193 /note= "HLA-binding peptide"
FT Region 194..213 /note= "HLA-binding peptide"
FT Region 214..233 /note= "HLA-binding peptide"
FT 248..267 /note= "HLA-binding peptide"
FT Region 289..308 /note= "HLA-binding peptide"
FT Region 306..325 /note= "HLA-binding peptide"
FT Region 311..330 /note= "HLA-binding peptide"
FT Region 349..368 /note= "HLA-binding peptide"
FT Region 365..384 /note= "HLA-binding peptide"
FT Region 383..402 /note= "HLA-binding peptide"
FT Region 403..422 /note= "HLA-binding peptide"
FT 403..422 /note= "HLA-binding peptide"
FT W02003075845-A2.
PN 18-SEP-2003.
XX
PD 07-MAR-2003; 2003WO-US006962.
XX
PF 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
DR GENBANK; AY046410.
DR
XX
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 6; Page 7; 49pp; English.
XX
XX
CC The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match 97.8%; Score 2070.5; DB 7; Length 431;
Best Local Similarity 98.6%; Pred. No. 9.5e-132;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLLSSKIVARNESLKQOARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAQLNLAELTTAQQOKATLVAQKAAAEAEARQAAAAQAAAAEA 240
DB 181 TTAQNTNALNTQOAELEAQLNLAELTTAQQOKATLVAQKAAAEAEARQAAAAQAAAAEA 240
QY 241 KAAAEAKALQEAQAQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGSDTDSQAQAQVNN 300
DB 241 KAAAEAKALQEAQAQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGSDTDSQAQAQVNN 299
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360
DB 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYVTVGVGGQIQVQEBANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYVTVGVGGQIQVQEBANYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYYPN 432
DB 420 FNPGSVSIYYPN 431

RESULT 8

ADX37272
ID ADX37272 standard; protein; 431 AA.
AC ADX37272;
DT 21-APR-2005 (first entry)
DE Streptococcus mutant glucan binding protein B variant #1.
XX immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus mutans.

US2005031633-A1.

10-FEB-2005.

09-MAR-2004; 2004US-00797821.

13-APR-1998; 98US-0081550P.

08-JAN-1999; 99US-0115142P.

12-APR-1999; 99US-0029004P.

07-MAR-2002; 2002US-0363209P.

08-AUG-2002; 2002US-0402483P.

07-MAR-2003; 2003US-00383930.

(SMIT/) SMITH D J.

(TAUB/) TAUBMAN M A.

Smith DJ, Taubman MA;

WPI; 2005-151644/16.

New composition comprising a fragment of a glucan binding protein-B

(GbpB) that binds to MHC class II protein, and a biocompatible

microparticle, useful for producing an antibody (claimed) for immunizing

mammals against dental caries.

Claim 3; SEQ ID NO 29; 73pp; English.

The invention relates to a composition comprising a fragment of a glucan

binding protein-B (GbpB) and a biocompatible microparticle, where the

CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 431 AA;

Query Match 97.8%; Score 2070.5; DB 9; Length 431;
Best Local Similarity 98.6%; Pred. No. 9.5e-132; Indels 1; Gaps 1;
Matches 426; Conservative 1; Mismatches 4;

QY 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLLSSKIVARNESLKQOARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAQLNLAELTTAQQOKATLVAQKAAAEAEARQAAAAQAAAAEA 240
DB 181 TTAQNTNALNTQOAELEAQLNLAELTTAQQOKATLVAQKAAAEAEARQAAAAQAAAAEA 240
QY 241 KAAAEAKALQEAQAQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGSDTDSQAQAQVNN 300
DB 241 KAAAEAKALQEAQAQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGSDTDSQAQAQVNN 299
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360
DB 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYVTVGVGGQIQVQEBANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYVTVGVGGQIQVQEBANYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYYPN 432
DB 420 FNPGSVSIYYPN 431

RESULT 9

AE91500
ID AEB91500 standard; protein; 431 AA.

XX AEB91500;

DT 20-OCT-2005 (first entry)

XX Microbial pathogen adhesin protein sequence, SEQ ID NO:210.

DE algorithm; adhesin; pharmaceutical; vaccine; drug screening;

KW bordetella pertussis infection; antibacterial; pneumonia;

KW antiinflammatory; respiratory-gen.; gastric ulcer; antitumor;

KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.

OS Streptococcus mutans.

PN WO2005076010-A2.

XX 18-AUG-2005.

PD 07-FEB-2005; 2005WO-IN000037.

XX 06-FEB-2004; 2004IN-DE000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL) COUNCIL SCI & IND RES SOUTH AFRICA.

PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;

XX

PI

XX WPI; 2005-597835/61.
XX Computational method for identifying adhesin and adhesin like molecules,
PT comprises computing sequence-based attributes of protein sequences using
PT neural network software and training an artificial neural network.
XX
XX Claim 16; SEQ ID NO 210; 402pp; English.
XX
XX The present invention relates to a computational method (M1) for
CC identifying adhesin and adhesin-like proteins, by computing the sequence-
CC based attributes of protein sequences using five attribute modules of a
CC neural network software, training an artificial neural network (ANN) for
CC each of the computed five attributes, and identifying the adhesin and
CC adhesin-like proteins having probability of being an adhesin (Pad) as
CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
CC genes encoding adhesin and adhesin-like proteins, having 105 fully
CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)
CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
CC proteins, of therapeutic potential, and identifying and short-listing
CC proteins for further testing in development of new vaccine formulations
CC to eliminate diseases caused by various pathogenic organisms. (M1) is
CC useful for identifying putative adhesins that are important in drug
CC discovery and preventing therapeutics for whooping cough, pneumonia, from
CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from
CC distantly related organisms, and from bacteria belonging to a wide
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
CC unique proteins. The present sequence is a microbial pathogen adhesin
CC protein sequence.
XX
XX Sequence 431 AA;
XX
XX Query Match 97.8%; Score 2070.5; DB 9; Length 431;
XX Best Local Similarity 98.6%; Pred. No. 9.5e-132;
XX Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQKATLVAKAAAEAAQAAAAQAAAA 240
DB 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQKATLVAKAAAEAAQAAAAQAAAA 240
QY 241 KAAAEKALQEAQAQAQAANNNTCAATDASQQAQAANTQCAQTGSDTDSQAQAVNN 300
DB 241 KAAAEKALQEAQAQAQAANNNTCAATDASQQAQAANTQCAQTGSDTDSQAQAVNN 299
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
DB 300 SQESTTTATEAPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
QY 361 ASAAAAGYRVGTPSPSAGAVVNDGGYGHVAVYTVGGQGIQVQEANYAGNOSIGNYRCW 420
DB 360 ASAAAAGYRVGTPSPSAGAVVNDGGYGHVAVYTVGGQGIQVQEANYAGNOSIGNYRCW 419
QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 10
ADD93653
ID ADD93653 standard; protein; 431 AA.
XX
XX ADD93653;
AC
XX 29-JAN-2004 (first entry)
DT
XX Streptococcus mutans glucan binding protein-B.
DE
XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
KW
XX Streptococcus mutans.
OS
XX WO2003075845-A2.
PN
XX 18-SEP-2003.
PD
XX 07-MAR-2003; 2003WO-US006962.
XX
XX 07-MAR-2002; 2002US-0363209P.
PR
XX 08-AUG-2002; 2002US-0402483P.
PR
XX (FORS-) FORSYTH INST.
PA
XX Smith DJ, Taubman MA;
XX
XX WPI; 2003-845091/78.
DR
XX GENBANK; AY046414.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 5; Page 8-9; 49pp; English.
XX
XX The present sequence is the protein sequence of the glucan binding
CC protein-B (GbpB) of Streptococcus mutans strain 5SM3. The sequence
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC invention provides immunogenic compositions and vaccines for dental
CC caries. The compositions comprise major histocompatibility complex (MHC)
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC covalently linked with peptide subunits of a glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Dieptopic or multieptopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
XX Sequence 431 AA;
XX
XX Query Match 97.7%; Score 2066.5; DB 7; Length 431;
XX Best Local Similarity 98.4%; Pred. No. 1.8e-131;
XX Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEXMLHQEQDKAAVEQKHQENQAANTVAANOE 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQKATLVAKAAAEAAQAAAAQAAAA 240
DB 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQKATLVAKAAAEAAQAAAAQAAAA 240

QY 241 KAAAEKALQEQAAQAAANNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 300
DB 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 299
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 360
DB 300 SQESTTTATAQPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 359
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 420
DB 360 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYPN 432
DB 420 FNPGSVSIYPN 431

RESULT 11
AD37276
ID ADX37276 standard; protein; 431 AA.
AC ADX37276;
DT 21-APR-2005 (first entry)
XX Streptococcus mutant glucan binding protein B variant #5.
DE immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus mutans.
OS
XX
XX US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
XX 08-JAN-1999; 99US-0115142P.
XX 12-APR-1999; 99US-00290049.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 33; 73pp; English.
PS
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 431 AA;

Query Match 97.7%; Score 2066.5; DB 9; Length 431;
Best Local Similarity 98.4%; Pred. No. 1.8e-131;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60

DB 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
QY 61 QVSALQTOQAELOAENQRLQEAQSAATLGGQIQTLSSKIVARNESLKQQAQSAQKSAQNTSY 120
DB 61 QVSALQTOQAELOAENQRLQEAQSAATLGGQIQTLSSKIVARNESLKQQAQSAQKSAQNTSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQAAVEQKQOENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQAAVEQKQOENQAAINTVAANQE 180
QY 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEBAFQAAAAQAAAA 240
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEBAFQAAAAQAAAA 240
QY 241 KAAAEKALQEQAAQAAANNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 300
DB 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 299
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 360
DB 300 SQESTTTATAQPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 359
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 420
DB 360 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYPN 432
DB 420 FNPGSVSIYPN 431

RESULT 12
ABP29684
ID ABP29684 standard; protein; 447 AA.
XX AC ABP29684;
XX DT 02-JUL-2002 (first entry)
XX Streptococcus polypeptide SEQ ID NO 8544.
DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW Streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettelin H;
XX WPI; 2002-352536/38.
DR N-PSDB; ABN70315.
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
PS Claim 1; Page 3965; 4525pp; English.
XX

CC The invention relates to a protein (ABP25413-ABP10895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I) may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 XX Sequence 447 AA;

Query Match 58.1%; Score 1229.5; DB 5; Length 447;
 Best Local Similarity 59.7%; Pred. No. 7.9e-75;
 Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;
 QY 1 MKKRILSAVLVSGVTLSATTLSSAVKADDFDAQIASQDSKINLTAQQAQAQVNTIIG 60
 DB 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSGQQAQAQVNTAIKG 58
 QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIOTLSKIVARNESLKQARSQAQSNATSY 120
 DB 59 QVGALESQOQSELEAQAQLEAVSQQLGQEIOTLSKNIVARNESLKQVRSQAQGN-LTNV 117
 QY 121 INAINSKSVSDAIRVSAIREVVSANEKMLHQEQDKAAVQKQHQENQAINTVAANOE 180
 DB 118 INTILNSKSVDAVNVVVAIREVVSANEKMLAQOEADKALEAKQIENQNAINTVAANKQ 177
 QY 181 TTAQNTNALNTQQAQLAAQLNLQALTLTAQOKATLVQAQAAAEARQAQAQAQAABA 240
 DB 178 ALENKAALATQRAQLAAQLSALQTLTVQNEKASLIQAKAQAEEAKAQAQAABA 237
 QY 241 KAAAEKALQEOAQA 300
 DB 238 KAQA 285
 QY 301 SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSANTYPAG 336
 DB 286 SATTVATTATATNEPKVTQSPVVTKAVEAPKAVVSTTPRAVSKPVVRSYDSSNTYPNG 345
 QY 337 QCTWGVKSLAPVWNGYNGGQMAASAAAAGYRVGSTPSAGAVAVW--NDGGYGHVAVVT 394
 DB 346 QCTWGAKSMAVWNGYNGNQGASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVVT 405
 QY 395 GV-QGGQIQVQEAANYAGNOSIGNYRWNP---GSVSYIYPN 432
 DB 406 SVANNSSIQVMSNYSAGNMSIGNYRGSFNPSPASGSVYIYPN 447

RESULT 13
 ADU69524
 ID ADU69524 standard; protein; 447 AA.
 XX
 AC ADU69524;
 XX
 DT 10-FEB-2005 (first entry)
 XX
 DE S agalactiae hyperimmune serum reactive antigen seqid 219.
 XX
 KW immune stimulation; antigen; bacterial surface display;
 KW hyperimmune serum reactive antigen; vaccine; bacterial infection;
 KW antibacterial; infection.
 XX
 OS Streptococcus agalactiae.
 XX

PN W02004099242-A2.
 XX 18-NOV-2004.
 XX 06-MAY-2004; 2004WO-EP004856.
 XX 07-MAY-2003; 2003EP-00450112.
 PR 28-NOV-2003; 2003EP-00450266.
 XX (INTE-) INTERCELL AG.
 XX Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;
 PI WPI; 2004-821662/81.
 DR N-PSDB; ADU69307.
 XX New nucleic acid molecule encoding a hyperimmune serum reactive antigen,
 PT useful for the manufacture of a vaccine against Streptococcus agalactiae
 PT infection.
 XX Claim 14; SEQ ID NO 219; 221pp; English.

CC The invention describes an isolated nucleic acid molecule encoding a
 CC hyperimmune serum reactive antigen or its fragment. Also described are: a
 CC vector comprising the nucleic acid molecule; a host cell comprising the
 CC vector; a hyperimmune serum-reactive antigen comprising a sequence
 CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467
 CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens
 CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;
 CC a process for producing a Streptococcus agalactiae hyperimmune serum
 CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical
 CC composition, especially a vaccine, comprising the hyperimmune serum-
 CC reactive antigen or nucleic acid molecule; an antibody that binds at
 CC least to a selective part of the hyperimmune serum-reactive antigen; a
 CC hybridoma cell line, which produces the antibody; a method for producing
 CC the antibody; an antagonist that binds to the hyperimmune serum-reactive
 CC antigen; a method for identifying an antagonist capable of binding to the
 CC hyperimmune serum-reactive antigen; a method for identifying an
 CC antagonist capable of reducing or inhibiting the interaction activity of
 CC a hyperimmune serum-reactive antigen to its interaction partner; a
 CC process for in vitro diagnosing a disease related to expression of the
 CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis
 CC of a bacterial infection, especially a S. agalactiae infection. The
 CC hyperimmune serum reactive antigen is useful for isolating, purifying
 CC and/or identifying an interaction partner of the hyperimmune serum
 CC reactive antigen. The hyperimmune serum reactive antigen is useful for
 CC generating a peptide binding to the hyperimmune serum reactive antigen,
 CC where the peptide comprises anticalines, or for the manufacture of a
 CC functional nucleic acid comprising aptamers or spiegelmers. The nucleic
 CC acid molecule is useful for the manufacture of a functional ribonucleic
 CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic
 CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful
 CC for the manufacture of a vaccine against S. agalactiae infection. This is
 CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum
 CC reactive antigen.

XX Sequence 447 AA;
 Query Match 58.1%; Score 1229.5; DB 8; Length 447;
 Best Local Similarity 59.7%; Pred. No. 7.9e-75;
 Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;
 QY 1 MKKRILSAVLVSGVTLSATTLSSAVKADDFDAQIASQDSKINLTAQQAQAQVNTIIG 60
 DB 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSGQQAQAQVNTAIKG 58
 QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIOTLSKIVARNESLKQARSQAQSNATSY 120
 DB 59 QVGALESQOQSELEAQAQLEAVSQQLGQEIOTLSKNIVARNESLKQVRSQAQGN-LTNV 117
 QY 121 INAINSKSVSDAIRVSAIREVVSANEKMLHQEQDKAAVQKQHQENQAINTVAANOE 180
 XX
 XX

Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQBELTTAQDQKATLVAKAAABEAAARQAAAAA 240
Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAAARKAAEAQAAEA 237
QY 241 KAAAEKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTDSQAQAQVNN 300
Db 238 KAAAEKALQEQAAQAAQAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285
QY 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336
Db 286 SSATTVATTATATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPNG 345
QY 337 QCTWGVKSLAPWVGNVGNWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVYT 394
Db 346 QCTWGAKSMAWVGNVGNWGNANQWGAASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQEQANYAGNOSIGNYRGWNP---GSVSYIYPN 432
Db 406 SVANNSSIQVMESNYAGNMSIGNYRGSFNPSAGSVYIYPN 447

RESULT 14
ADV81808
ID ADV81808 standard; protein; 447 AA.
AC ADV81808;
DT 24-FEB-2005 (first entry)
DE Streptococcus agalactiae protein sequence, SEQ ID 786.
XX Antibacterial; vaccine; bacterial infection.
KW Streptococcus agalactiae.
OS Streptococcus agalactiae.
PN FR2824074-A1.
XX 31-OCT-2002.
XX 26-APR-2001; 2001PR-00005642.
XX 26-APR-2001; 2001PR-00005642.
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
PI Glaser P, Rueniok C, Chevallier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst P;
XX WPI; 2004-101891/11.

Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.
Claim 6; SEQ ID NO 786; 2687pp; French.
The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II; ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. CC Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2 contains 6617 sequence whereas the present patent only contains 2344 sequences.
XX
SQ Sequence 447 AA;
Query Match 58.1%; Score 1229.5; DB 8; Length 447;
Best Local Similarity 59.7%; Pred No. 7.9e-75;
Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;
QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTI 60
Db 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFOSKIATDSVINTLSGQQAQAQVNTAIKG 58
QY 61 QVSALQTOQAELQAEQLEAQSATLGOQIOTLSSKIIVARNESLQQAARSQAQNAATSY 120
Db 59 QVGALESQCSLEAQAQLEAVSQQLGOEIQTLNKKIIVARNESLKKQVRSQAQGN-LTNY 117
QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQODKAAVEOKHOEQAAINTVAANO 180
Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQBELTTAQDQKATLVAKAAABEAAARQAAAAA 240
Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAAARKAAEAQAAEA 237
QY 241 KAAAEKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTDSQAQAQVNN 300
Db 238 KAAAEKALQEQAAQAAQAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285
QY 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336
Db 286 SSATTVATTATATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPNG 345
QY 337 QCTWGVKSLAPWVGNVGNWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVYT 394
Db 346 QCTWGAKSMAWVGNVGNWGNANQWGAASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQEQANYAGNOSIGNYRGWNP---GSVSYIYPN 432
Db 406 SVANNSSIQVMESNYAGNMSIGNYRGSFNPSAGSVYIYPN 447

RESULT 15
ADV81808
ID ADV81808 standard; protein; 447 AA.
AC ADV81808;
DT 24-FEB-2005 (first entry)
DE Streptococcus agalactiae protein, SEQ ID 2949.
XX Antibacterial; vaccine; bacterial infection.
KW Streptococcus agalactiae.
OS Streptococcus agalactiae.
XX WO200292818-A2.
XX 21-NOV-2002.
XX 26-APR-2002; 2002WO-IB003059.
XX 26-APR-2001; 2001PR-00005642.
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
PI Glaser P, Rueniok C, Chevallier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.

PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.

XX
PS
XX
XX Claim 6; SEQ ID NO 2949; 439pp; French.

CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (i) or (ii) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.

XX Sequence 447 AA;

Query Match		58.1%;	Score 1229.5;	DB 8;	Length 447;
Best Local Similarity		59.7%;	Pred. No. 7.9e-75;		
Matches 276;		Conservative 42;	Mismatches 99;	Indels 45;	Gaps 9;
QY	1	MKKRILSAVLVSGVTLSSTATTISAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60		
Db	1	MKKRILSAVLVSGVTILGTAAV--TVNADDFDSKIAATDSVINTLSQQAQAQVNTAIKG	58		
QY	61	QVSALQTQQAELQAEQNRLEAQSATILGQOIOTLSSKIVARNESLKKOAESAKSNAATSY	120		
Db	59	QVGALESQOSELEAQAQAEVSVQQLGQEIQTLSNKKIVARNESLKKQVRSQKGN-LTNY	117		
QY	121	INAIINSKSVSDAIRVSVANERKMLHQEQDKAAVEQKHQENQAAINTVAANOE	180		
Db	118	INTILNSKSVSDAVNRVVAIRVVSVANERKMLAQEQEADKALEAKQIENQNAINTVAANKQ	177		
QY	181	TTAQNNTALNTQQAQLEAAQLNLQAEELTTAQDOKATLVAKAAAEAEARQAAAAAQAEEA	240		
Db	178	AIENNKAAALATQRAQLEAAQLSALQTTVQNEKASLIQAKAQAEAEARQAAAAAQAEEA	237		
QY	241	KAAAEKALQEQQAQAAQAAANNNTQATDASDQQAADNTQAACTGDSDDQSAQAQVNN	300		
Db	238	KAAQAEAKAQAESVAKAQAAA-----QVESATAPTETVQTPTPTKPSNLTAT	285		
QY	301	SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSANTYPAG	336		
Db	286	SSATTVAITATATNEPKVTQPSVTVKAEAPKAVSVSTPRAVSKPVRSYDSSNTYPNG	345		
QY	337	QCTWGVKSLAPVGNVWNGGQWAAASAAAAYRVGSTPSAGAVVW--NDGGYGHVAVYT	394		
Db	346	QCTWGAKSMAWVGNVWGNQWGAASRAAGYSGVTTPRVGAVAVVWPDGCGYGHVAVYT	405		
QY	395	GV-QGGQIQVQEAQYAGNQSIGNYRGWNP---GSVSYIYPN	432		
Db	406	SVANNSSIQWMSNYSVAGNMSIGNYRGSFNPSASGSVYIYPN	447		

Search completed: February 15, 2006, 18:03:00
Job time : 114.754 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:03:21 ; Search time 21.0162 Seconds
(without alignments)
1977.789 Million cell updates/sec

Title: US-10-797-821-32
Perfect score: 2116
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGWNFGSVSYIYPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	49.8	392	2 G95258	secreted 45 kd pro
2	1053	49.8	392	2 B98124	general stress pro
3	681.5	32.2	461	2 JN0097	secreted 45K prote
4	658	31.1	456	2 E86903	hypothetical prote
5	537.5	25.4	507	2 S05542	hypothetical prote
6	356	16.8	398	2 AC1763	peptidoglycan lyti
7	343.5	16.2	401	2 A11387	cell wall-binding
8	319	15.1	473	2 F70031	conserved hypothet
9	259.5	12.3	581	2 E75383	cell wall-binding
10	256	12.1	461	2 H84099	cell wall binding
11	254.5	12.0	436	2 AH1387	cell wall binding
12	251	11.9	437	2 AB1763	conserved hypothet
13	237	11.2	528	2 B75310	tolA protein - Eac
14	230	10.9	421	2 JV0057	probable tail fibe
15	228	10.8	971	2 B90835	probable membrane
16	228	10.8	973	2 C85693	hypothetical prote
17	226.5	10.7	255	2 G90061	hypothetical prote
18	226.5	10.7	265	2 B9837	hypothetical prote
19	226.5	10.7	267	2 F90028	hypothetical prote
20	226.5	10.7	394	2 F90725	membrane spanning
21	226.5	10.7	394	2 G85576	membrane spanning
22	225.5	10.7	166	2 G90029	hypothetical prote
23	225.5	10.7	392	2 F96937	cell wall-associat
24	222	10.5	492	2 A28616	M5 protein precurs
25	221.5	10.5	688	2 A83179	conserved hypothet
26	221	10.4	1122	2 G64887	probable tail fibe
27	217	10.3	1528	2 A60338	surface antigen A
28	217	10.3	4776	2 E95206	cell wall surface
29	215	10.2	539	2 A28549	M24 protein precur

RESULT 1
G95258
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: G95258
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
on, J.D.; Unyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: G95000; PMID:21357209; PMID:11463916
A:Accession: G95258
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-392 <KUR>
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AB005672; I
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2216

Query Match 49.8%; Score 1053; DB 2; Length 392;
Best Local Similarity 49.8%; Pred. No. 2.3e-45;
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;

QY	1	MKKRILSAVLVSGVTLSATILSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60
DB	1	MKKKILASLLSTVMVSVQAVLTAAHAEITDDKIAAQDNKISNLTAAQQAQAQVDQIQE	60
QY	61	QVSALQTOQAELQAEENORLEAQSATLGOQIOTLSSKIVARNESLKOQARSQAKSNAATSY	120
DB	61	QVSAIQAEQSNLQAEENORLEAQSATLGOQIOTLSSKIVARNESLKOQARSQAKSNAATSY	120
QY	121	INAIINSSVSDAINRVSAIREVWSANFKMLHQEQDQKAAVEQKHQHOQAATNTVAANQE	180
DB	121	INTIVNSKITEAISRVAAVMEIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ	180
QY	181	TTAQTNTALNTCOAQLAAQNLQAEELTAQDOKATLVAQKAAAEAAQAAQAQAAEA	240
DB	181	KLADDAQALTTKQAEKAAELSLAAEKATAGEKASLLEQKAAAEAAQAAQAQAAEA	240
QY	241	KAAAEKALQEAQAQAQAANNTNTQATDASDQQAADNDTQAAOTGDSTQDQAQAQVNN	300
DB	241	KRAQQQSV-----LASANTLTAQQAQVSESAAAQVRAKVRPT-----	279
QY	301	SDQESTTATAAQAQPSASSASTAAVAANTSSANTYPAGQCQTGWVKSLAPVGNVWNGGQWA	360
DB	280	-----YSTNASSYPIGECTGWVKTLAPWAGDYGNGAQWA	314
QY	361	ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVTVTQVG-QGQVQEAQVAGNQSTGNVYG	419

ALIGNMENTS


```
Db 1 MKKKIISAILMTSTVVLSSAAPISGVVYADT-NSDIAKQDATISSAOSAKAQAQAQVDSLQS 59
QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQOARSQAQKSNAAATSY 120
Db 60 KVDLSLOQKQASTYAKIAKIESEAKALNAQIATLINESIAERTKTLEQAARSAQVNSATNY 119
QY 121 INAIINSSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVQKQHQENQAAINTVAANOQE 180
Db 120 MDVAVNSKSLTDVIQKVTAIATVSSANKQMLEQEQKEQKELSKQSTVKYKNTYNQFVLSLQ 179
QY 181 FTAQNTNALNTQOALEAAQINLQALTTAODOKATLVAQKAAAEAPQAAAAQAAAAEA 240
Db 180 SLDSQAQELTSQAELKVATLNYQATIAQDKKOSLDEKAAAEKAAQEAQKQAAYEA 239
QY 241 KAAAEAKALQEQAQAQAANNNTNTQATDASDQAAAAADNTQAAQTGDSTDQSAQA--AV 298
Db 240 Q-----QKEAQAQAATAATYKAV---BEATSTVSSSQASQSSSSNTSSNTSS 288
QY 299 NNSDQESTTATAAQPSSASSASTAAVAANTSS-----ANTYPAG 336
Db 289 NSSSSSSSSSSSSSSSSSSSGGSGTNTGNNAAGTGTGSSSGINGSTPIANPYAGG 348
QY 337 QCT---WG-----VKSLAPWVGNWCGCGOWAASAAAAG--YRVGSTPAGAVAV-- 381
Db 349 GCTDYWQYFAAQGYIRIMP-----GNGGOWATNGPAQGVHLHVGAAP--GVIASSF 400
QY 382 -----WMDGGYGHVAYVTGV-QGGQIOVOEANYAGNQSIGNVRGFWNPGSVSYTYPN 432
Db 401 SADPFGVYANSPGHVAIVKSVNSNGTITIKESGY-GTTWGHGR-TVVSAGVTFLMPN 456

RESULT 5
S05542
hypothetical protein, 54K - Enterococcus faecium
C:Species: Enterococcus faecium
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S05542
R:Fuerst, P.; Moesch, H.U.; Solioz, M.
Nucleic Acids Res. 17, 6724, 1989
A:Title: A protein of unusual composition from Enterococcus faecium.
A:Reference number: S05542; MUID:89385998; PMID:2780297
A:Accession: S05542
A:Molecule type: DNA
A:Residues: 1-507 <FUE>
A:Cross-references: UNIPROT:PI3692; UNIPARC:UPI000016f6FC; GB:X16421; EMBL:M26048; NID:9
A>Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 25.4%; Score 537.5; DB 2; Length 507;
Best Local Similarity 28.7%; Pred. No. 8.7e-20;
Matches 146; Conservative 89; Mismatches 172; Indels 101; Gaps 12;

QY 10 LVSGVTLSSATLSAVKADDFDAQIASODSKINNLTAAQOAAQAQVNTTQGOVSALOTQ 69
Db 1 MLSSIALTAVGSPIAAAADDFSQIQQDKKIADLQNOQAASQOIEALEGGVSAINTKA 60
QY 70 ASLQENORLEAQSATLGOQIOTLSSKIVARNESLKQOARSQAQKSNAAATSYNAIINSS 129
Db 61 QDLLTKQDTLRKESQLKQEIKDQERIEKRAITQKQARETOVKNTSSNYDAVLNAD 120
QY 130 VSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKQHQENQAAINTVAANOQETIAQNTNAL 189
Db 121 LADAVGRIQAMSTIVKANQDLVQQKEDQKQAEAKKAENEAKQKELADNQAALQKQKGL 180
QY 190 NTQOALEAAQINLQALTTAODOKATLVAQKA-----AAEEAARQAARQAQ 235
Db 181 LAKQADNLVLTSLAAEQATSEDKKADLNRKKAEEAEQARIREQARLAEQARQAQAEK 240
QY 236 AAEEAKAAAEAKALQEQAQAQAQAANNNTNTQATDASDQAAA-----AADNTQ 282
Db 241 AEKAREQAQAQAQATQALSSASTTTESSAAQSSEESKAPESSTTEBTESTESTTTE 300
QY 283 AAQTGDSTDQSA-----AQAVNNSDQESTTATAAQPSSASSASTAAVAANTSSANT----- 332
```

```
Db 301 NSSTGSSSTESSTESTVPESTQESTPANTPESSSSSNTNNTNNTNNTNNTNNTNST 360
QY 333 -----YPAGQCTMGVKSLA-----PMWGNVWNG----- 356
Db 361 NNNNNNTVTPAPTPTTPAPAPAPNPSGVNVAIAVEAYKYIGTPYV---WGGKDPG 417
QY 357 -----GOWAASAAAAGYRVG-STPSAGAVAVVND-GGIGHVAYVTG 395
Db 418 FDCSGTRVYVLQVTRDIGMTVPQESAGTKISVSQAKAGDLLFWGSAGTYHVAISLG 477
QY 396 VGGQIQVQEANYAGNQSIGNVRGWFNP 423
Db 478 --GGQ-YIHAPQGENVKGVQ-WYTP 501

RESULT 6
AC1763
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clp11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1763
R:Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1763
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <GLA>
A:Cross-references: UNIPROT:O927Y8; UNIPARC:UPI00000CC935; GB:AL592023; PIDN:CAC97875.1;
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: spl

Query Match 16.8%; Score 356; DB 2; Length 398;
Best Local Similarity 27.1%; Pred. No. 6.2e-11;
Matches 114; Conservative 87; Mismatches 147; Indels 72; Gaps 16;

QY 1 MKKRILSAV-LVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTQ 59
Db 1 MKKNTFIAISLAAVISLAPFTTNVFA--DVNTDIQNDKKINDIKSKKTKGLQSLSSLV 58
QY 60 GOVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQOARSQAQKSNAAAT 119
Db 59 ADLEKAQEKAKSLOGEPDQTKELQNLQDQIKDINERIKERETVLKERARAWQKTSNSNA 118
QY 120 YINAIINSSVSDAINRVSAIREVVSANEKMLHQEQD-----KAAVEQKQHQENQAAI 172
Db 119 YLEVLIDRAENLSDLVGRVSAVNVQLVDSKLSILEDQKDEKALKTKQTAVKKQEQAATAI 178
QY 173 NTVAANOQETIAQNTNALNTQOALEAAQINLQALTTAODOKATLVAQKAAAEAAARQAA 232
Db 179 HEFEAQ-----NKIEAQKAEKAIVAQLAADQSAENKAGLVSE--DKAAKE-- 226
QY 233 AAQAAAEKAAAEAKALQEQAQAQAQAANNNTNTQATDASDQAAAADNTQAAQTGDSTDQ 292
Db 227 -----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTS 258
QY 293 SAAQAVNNSDQESTTATAAQPSSASSASTAAVAANTSSANTYPAG-----OCTMCKVSL 345
Db 259 NKVESTNNSEAPS-----HATPSGGGYSAMIAARAQLGKPYSLGATGPSAFPDCS-GFTSY 313
QY 346 ---APWVGNWNGCGOWAASAAAAGYRVGSTPSAGAVAVN-DGGYGHVAYVTGQGGQI 401
Db 314 APRAAGVSLPRTSGGQYAAASKIS-----ASQAKPGDLVFFNYGGGIAHVGIVVG--GGQM 367

RESULT 7
```

Al1387
peptidoglycan lytic protein p45 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: Al1387
R:Glaser, P.; Prangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: Al1387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <GLA>
A:Cross-references: UNIPROT:Q9RE04; UNIPARC:UPI00000D019D; GB:NC_003210; PIDN:CAD00583.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: spi

Query Match 16.2%; Score 343.5; DB 2; Length 401;
Best Local Similarity 26.4%; Pred. No. 2.6e-10;
Matches 111; Conservative 92; Mismatches 148; Indels 69; Gaps 15;
QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQ 59
DB 1 MKKNTFIAISLAAVISLTPAFTTNVFA--DVNTDIQNQDKINDIKSKKTDLQSLGLIV 58
QY 60 GQVSALQTOQAELQENORLEAQSATLGOIOTLSKTVARNESLKOQARSQAQKSNAA 119
DB 59 ADLEAQAQKAKLQGEFKTKGELKLNEDIKINERIKERTVLKERARAKQTSNSNA 118
QY 120 YINAIINSKVSDAINRVSASREVVSANERKMLHQEQDKAAVE-----QKHQENQA-AI 172
DB 119 YLEVLDAENLSDLGRVSNVQLVSDSKS ILEDQONDEKALKTKQTAVKKKQEQDQATAI 178
QY 173 NTVAAQETIAQNTALNTQQAQLEAQLNLAELTTAQQOKATVAQKAAEEAARQAA 232
DB 179 HEYEAQQ-----NKIEQAQKAEKAI VAQLASDAQSAENAKAGLVSE--DKAKE-- 226
QY 233 AAQAAAEAKAAEAKALQEOAAQAAQAAANNNTQATDASDAQAAAAAATQAAQTGDSTDQ 292
DB 227 -----ATARATALRE-----ATSSNVGSESSDSTPTSKSNTTK 262
QY 293 SAAQAVNNSDQBSTTATAAQAQSSASTAAVAANTSSANTYPAG-----QCTWGVKSL 345
DB 263 NVASNDNS-----APSAATPSSGGYSAMI SAANAQLGPKYSLGATGPSAFDCS-GFTSY 316
QY 346 ---APWVNYWNGGQWAAASAAAAGRYVGSTPSAGAVAVWNGD-GYGHVAVYVGQGOI 401
DB 317 AFRAAGVSLPRTSGQGYAAASKIS----ASQAKPGDLVFPFNYGSGIAHVGIYVG--GGQM 370

RESULT 8
F70031
cell wall-binding protein homolog yvcE - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: F70031
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Geller, iech, J.; Harwood, C.R.; Henaat, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F. Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maguda, S.; Maueel Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A:Authors: Schleich, S.; Schroeter, F.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: F70031
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-473 <KUN>
A:Cross-references: UNIPROT:P40767; UNIPARC:UPI0000060AB1; GB:Z99121; GB:AL009126; NID:9 A:Experimental source: strain 168
C:Genetics:
A:Gene: yvcE

Query Match 15.1%; Score 319; DB 2; Length 473;
Best Local Similarity 23.9%; Pred. No. 5.1e-09;
Matches 111; Conservative 94; Mismatches 187; Indels 72; Gaps 11;
QY 1 MKKRILSAVLVSGVTLS-----ATTL SAVKADDFDAQIASQDSKINNLTAAQOAAQAAQ 54
DB 1 MKKSLITLGLASVIGTSSFLIPFTSKTASAEFLDEKKQKIESQSEV---ASSIEAKEKE 57
QY 55 VNTIQGVSAQTQAEL-----QAENORLEAQSATLGOIOTLSKIVAR 100
DB 58 LTELQENQSKIEKELKNDKALDTSNKIEDKKEENDTKKEIKLKEIKETEAREIEKR 117
QY 101 NESLKOQARSQAQKSNAAATSYINAIINSKVSDAINRVSASREVVSANERKMLHQEQDKAA 160
DB 118 NEILKRVRSIQESGGSGYIDVLGTSFGDFISRATVSSIVDADKDLIKQQEQDKAK 177
QY 161 VE-----QKHQENQAANTVAANQETIAQNTALNTQQAQLEAQLNLAELTTAQQD 213
DB 178 LEDSEADLNDKLKEVQAALAKLETWQKDLQNLNEKDLFDEAKASQKTKAKAISLKE 237
QY 214 KATLVAQKAAEEAARQAAQAAEAKAAEAKALQEOAAQAAQAAANNNTQATDASQ 273
DB 238 ASELANKANTE-----AEQARIKKEQEAALAIKKQBEAQKA-----SDETQTDSD 284
QY 274 QAAADNTQAAQTGDSTDSAAQAVNNSDQSTTTATAAQPSSASSASTAAVAANTSSANTY 333
DB 285 QTATTESSKASDDSDSDNSDSSNGSSNGSSNGSSNGSSNGSGTVISNSGGIE 344
QY 334 PAGQCTWGVKSLAPVWVNYWNGG-----QWA-ASAAAAGRYVGSTPS 375
DB 345 GAISVSGSIVGQSP----YKFGGRTQSDINNRI FDCSFFVRWAYASAGVNLGPVGTTT 400
QY 376 AGAVAVWNGG YGHVAVYVGQGOIQVQEAANYAGNQSIGNYRG 419
DB 401 DTLV-----GRGQAVSASEMKRGDL-VFFDTYKTNGHVGIYLG 437

RESULT 9
E75383
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: E75383
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; i, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.W. Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75383
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-581 <WHI>
A:Cross-references: UNIPROT:Q9RU45; UNIPARC:UPI00000C195A; GB:AE001998; GB:AE000513; NID: A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1549
A:Map position: 1
Query Match 12.3%; Score 259.5; DB 2; Length 581;

[illegible]

Db 282 -----AVVQPPSSSNEATETVSSGGGQFIKPASGILTSGFSERTN 322

or

5

109 BILBESIAERKDELLNDKRAKAMUQNGSISDTHIEVIDGHNSTQDFBKVVSADUSVIA 182

QY 151 LHQQEQDKAAVEQKHOENQAATNTVAANQETIAQNTNALNTQQAOLAAQINLQAEITTTA 210

QY 347 PWGNYWGN-----GGWAASAAAGYRVCGSTPSAGAVWVNDGGYCHVAVYVTGVGG 399
DB 323 PVTGKYSHKGODIAGGGTTVTVAASAGTVFSGFGASGSGF---GGYGVVVKIDHNGFG 379
QY 400 Q 400
DB 380 Q 380

RESULT 12
AB1763
cell wall binding protein homolog lin2647 [imported] - Listeria innocua (strain Clip1126
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AB1763
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreitt, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1763
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-437 <GLA>
A:Cross-references: UNIPROT:Q92Y79; UNIPARC:UPI0000CC934; GB:AL592022; PIDN:CAC97874.1;
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2647

Query Match 11.9%; Score 251; DB 2; Length 437;
Best Local Similarity 23.2%; Pred. No. 1.1e-05;
Matches 101; Conservative 80; Mismatches 149; Indels 106; Gaps 16;
QY 6 LSAVLVSGVTLSATTLSAVKAD-----DFAQIASQDSKINNLTAAQ 47
DB 11 LSLIIIS-----APLTSVKAESINDMKROSEIEQKKSELNKLNDTKNSLNHLENAE 63
QY 48 QAAA---QAOVNTIQGVSAQTQQAELQAEORLEAQSATLGQIQITLSSKIVARNESL 104
DB 64 KDAKELESLSIDETNKKLKEQEDKVDSENEKL-----KLKKEIEKLNDIRQKVL 119
QY 105 KQARSQAQNSAATSYINAINTSKVSADINRVSAIREVVSANERKMLHQEQDK---AAV 161
DB 120 DSRARAIQTGTATSYLDMPEADDFKELIDRTVTVSAIVKADQNMQDKDQKQKVA 179
QY 162 EKHQENQAAINTVAANOETIAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAQK 221
DB 180 ENSSEKLENLKVLAVELEVSNNMESQKKNLVMALAN-KDILT--KSEQTILLTNEQ 236
QY 222 AAEEAARQAAAQAAAKAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAADNT 281
DB 237 GALSDEQKLLANLAGEKAKQEAATKAAEEKMQEAAKN----- 276
QY 282 QAAQTGSDTDSAAQAQVNNSDQESTTATAQPSA--SSASTAAVAANTSSANTY--PA-G 336
DB 277 -----ATVAAQQSPSVTSAGTGATDTVSSGGQFIKPASG 312
QY 337 QCTWGVKSLA-PWGNVYGN-----GGWAAASAAAAGYRV-----GSTPSAGAVAVND 384
DB 313 MLTSGFSERTNPVTGKYSHKGQDIAGGGTITVSAASGRVVFSFGATGS-----GF 365
QY 385 GGYGHVAVTVGGQ 400
DB 366 GGYGVVVKIDHNGFGQ 381

RESULT 13
B75310
conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: B75310
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I
; Shen, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, M.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75310
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <WHI>
A:Cross-references: UNIPROT:Q9BSJ1; UNIPARC:UPI00000C1A76; GB:AE000513; NID:
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2133
A:Map position: 1

Query Match 11.2%; Score 237; DB 2; Length 528;
Best Local Similarity 29.0%; Pred. No. 6.4e-05;
Matches 94; Conservative 54; Mismatches 126; Indels 50; Gaps 14;
QY 5 ILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQGVSA 64
DB 48 VLSGMAISAASLAAPLLL-----NRNAVSTIAQADQLRPQIEALRTEVGTVOGELRA 99
QY 65 LQTOQAELQAEORLEAQSATLGQIQITLSSKIVARNESL-----KQARSQAQNSAATSY 120
DB 100 ARTER-----EAARSEAQKA--GQREAAROELAARQNLASAQEQEARLTQKQADLOQTR 152
QY 121 INAINTSKVSADINRVSAIREVVSANERKMLHQEQDKAAVEQKHQENQAAINTVAANO 180
DB 153 LKTUAEQRRLQEA--QAQAREKQLQASQKQL--QASEDRATQLDSQVLDLKURSAQAEQ 208
QY 181 TIAQNTNA-LNTQQAQLEAAQLNLQAEITTAQ--DQKATLVAQK-----AAAEAAARQAA 232
DB 209 --AQNAQTFANAQAQARTTELQRRRAAAQATAQAQATRAAQAQASQAQASARAQVREQAR 266
QY 233 AQAAAEAKAAAEKALQEQAAQAAQAAANNNTQATDASDQAAADNTQAAATGSDTDQ 292
DB 267 QAQRRAE---QAQARAQEQV-AQAQAAA-----QASVROQAQAAQTQLGQVRTGAAQ 314
QY 293 SAAQAQVNNSDQESTTATAQPSAS 316
DB 315 QVAAQ-----QQAQAQAHRFSSA 333

RESULT 14
JY0057
tolA protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: JY0057; B64810
R:Levengood, S.K.; Webster, R.E.
J. Bacteriol. 171, 6600-6609, 1989
A:Title: Nucleotide sequences of the tolA and tolB genes and localization of their produc
A:Reference number: JY0057; MUID:90078104; PMID:2687247
A:Accession: JY0057
A:Molecule type: DNA
A:Residues: 1-421 <LEV>
A:Cross-references: UNIPROT:P19334; UNIPARC:UPI0000137105; GB:M28232; NID:g148018; PIDN:f
A:Experimental source: strain JM105
A:Note: The authors translated the initiation codon GTG for residue 1 as Val
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B64810
A>Status: nucleic acid
A:Molecule type: DNA
A:Residues: 1-421 <BLAT>

A;Cross-references: UNIPARC:UPI0000137105; GB:AE000177; GB:U00096; NID:gl786955; PIDN:AA
A;Experimental source: strain K-12, substrain MG1655
C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach b
C;Genetics:
A;Gene: tolA
A;Map position: 17 min
A;Start codon: GTG
C;Keywords: nucleotide binding; P-loop; transmembrane protein
F;14-34/Domain: transmembrane #status predicted <MSS>
F;78-301/Domain: helical #status predicted <HGR>
F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 10.9%; Score 230; DB 2; Length 421;
Best Local Similarity 28.5%; Pred. No. 0.00011;
Matches 117; Conservative 65; Mismatches 179; Indels 50; Gaps 19;

QY 5 ILSAVLVSGVTLSSATTLISAVKADDFDAQIASQ-----DSKINNLTAQQQAQAQVNTIQ 59
DB 16 IISAVL--HVILFAALIWS-----FDENIEASAGGGSSIDA VVWDSGAVVEQVKRMQ 68
QY 60 GOVSA-----LQTOQA--ELQ-----AENQL- EAQSATLGGQIQTLSSKIVARNES 103
DB 69 SQESSAKSDEQKMKQQAABELEKQAEQERLQLEKERLAAQEQKQAEAKQAE 128
QY 104 LKQ-QARSAQSKNAATSYINAIINSKVSDDAENRVSAIREVVVANEKMLHQEQDQAAVE 162
DB 129 LKQQAEEAAAKAAADAKAEADAKAEAEAKK-----AAADAKKAEAEAAKAAAE 181
QY 163 -QKHENQAAI--NTVAANQETIAQNTNALNTQQAQLEAAQLNLAEL-TTAQDQKATLV 218
DB 182 AOKKAEAAALUKKAEAAEAARAAAEARKKAETAEAEKAEKAAAEKAAADKKA--A 239
QY 219 AQAQAAE-EAARQAAAAAQAQAAAEAKAAAEAKALQEQAQAQAQAANNNNT--QATDASDQA 275
DB 240 AEKAAADKKAQAEAAEAADKKAQAAAEKAAADKKAQAAAEKAAAEAAAEADDIFG 299
QY 276 AAADNTQAAQTDGSDTQSAQAQAVNNSDOESTTATTAQ-PSASSASTAAVAANTSSANTYP 334
DB 300 ELSSGKNAPKTTGGGAKGNASPAAGSNTKNGGASGADINNYAGQIKSAIESKFYDASSYA 359
QY 335 AGCCTWGVKSLAP--WVGNVWNGGOWA--ASAAAGYRVGSTPSAGAVAVW 382
DB 360 GKTCTLRIR-LAPDGMILLDIKPEGGDPAALCQAALAAAKLAKIPKPPSQAVY 409

RESULT 15
B90835
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: B90835
R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A39629; PMID:21156231; PMID:11258796
A;Accession: B90835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971 <HAY>
A;Cross-references: UNIPROT:Q8XQD4; UNIPARC:UPI000016542B; GB:BA000007; PIDN:BA035073.1;
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: EC01650

Query Match 10.8%; Score 228; DB 2; Length 971;
Best Local Similarity 26.4%; Pred. No. 0.00034;
Matches 96; Conservative 65; Mismatches 159; Indels 44; Gaps 10;

QY 19 ATTLISAVKADDFDAQIASQDSKINNLTAQQQAQAQVNTIQGVSAALQTOQAQAEQENOR 78
DB 121 AQNTAAKKSADASTAREAAATHATDAADSARAAASTSAGQAASSASAGTASTKA 180

Search completed: February 15, 2006, 18:10:24
Job time : 22.0162 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:21 ; Search time 123.762 Seconds
(without alignments)
2462.693 Million cell updates/sec

Title: US-10-797-821-32
Perfect score: 2116
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGMFNPQSVIYIPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2116	100.0	432	2	Q938V1_STRMU	Q938v1 streptococc
2	2099	99.2	432	2	Q938V2_STRMU	Q938v2 streptococc
3	2074.5	98.0	431	2	Q9AG98_STRMU	Q9ag98 streptococc
4	2070.5	97.8	431	2	Q938V3_STRMU	Q938v3 streptococc
5	2070.5	97.8	431	2	Q8DWM3_STRMU	Q8dwm3 streptococc
6	2066.5	97.7	431	2	Q938V0_STRMU	Q938v0 streptococc
7	1229.5	58.1	447	2	Q9AKA4_STRAG	Q9aka4 streptococc
8	1229.5	58.1	447	2	Q8E2H1_STRAS	Q8e2h1 streptococc
9	1229.5	58.1	447	2	Q8E7X9_STRAS	Q8e7x9 streptococc
10	1087.5	51.4	474	2	Q5M6K4_STRT2	Q5m6k4 streptococc
11	1087	51.4	398	2	Q5XEL1_STRP6	Q5xel1 streptococc
12	1087	51.4	398	2	Q7CNQ7_STRP8	Q7cnq7 streptococc
13	1087	51.4	398	2	Q8F3I8_STRP3	Q8f3i8 streptococc
14	1087	51.4	485	2	Q5M212_STRT1	Q5m212 streptococc
15	1081	51.1	398	2	Q9A1Z8_STRPY	Q9a1z8 streptococc
16	1053	49.8	392	2	Q8DMY4_STRR6	Q8dmv4 streptococc
17	1053	49.8	392	2	Q97N55_STRPN	Q97n55 streptococc
18	937	44.3	211	2	Q9ZAS7_STRMU	Q9zas7 streptococc
19	674.5	31.9	461	1	USF45_IACLC	P22865 lactococcu
20	658	31.1	456	2	Q9CDJ1_LACLA	Q9cdj1 lactococcu
21	585.5	27.7	524	2	Q9K2J9_ENTFC	Q9k2j9 enterococcu
22	569.5	26.9	516	1	P54_ENTFC	P13692 enterococcu
23	549	25.9	482	2	Q5M5M6_STRT2	Q5m5m6 streptococc
24	543.5	25.7	576	2	Q9KJ33_ENTHR	Q9kjj3 enterococcu
25	542	25.6	470	2	Q8QE3_9LACT	Q8rqe3 lactococcu
26	530	25.0	482	2	Q93LK4_ENTFA	Q93lk4 enterococcu
27	507.5	24.0	461	2	Q56SA7_STRTR	Q56sa7 streptococc
28	498.5	23.6	449	2	Q93LK3_ENTFA	Q93lk3 enterococcu
29	436.5	20.6	211	2	Q8DVU8_STRMU	Q8drv8 streptococc
30	416	19.7	226	2	Q93RG6_STRIT	Q93rg6 streptococc
31	415	19.6	544	2	Q840X3_STRMU	Q840x3 streptococc

ALIGNMENTS

RESULT 1
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1.
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SNI;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SNI;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match Best Local Similarity 100.0%; Score 2116; DB 2; Length 432;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKKRILSAVLVSGVTLSAT	TSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60
DB	1	MKKRILSAVLVSGVTLSAT	TSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60
QY	61	QVSAIQTOAEALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY	12	
DB	61	QVSAIQTOAEALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY	12	
QY	121	INAIINSKSVSDAIRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAATNTVAANQE	18	
DB	121	INAIINSKSVSDAIRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAATNTVAANQE	18	
QY	181	TTAQTNTALNTQQAQLEAAQLNLQALTTAQDQKATLVAQKAAAEAEARQAQAAQAAAEAA	24	

```
Db      181  TTAQNTNALNTQQAQLEAAQLNLQAELETTAQQKATLVAKKAAAEAAARQAAAAQAAAEA 240
Qy      241  KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
Db      241  KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
Qy      301  SQESTTATAAOPSPASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVNTWNGGOWA 360
Db      301  SQESTTATAAOPSPASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVNTWNGGOWA 360
Qy      361  ASAAAAGYRVGTPSPAGAVAVVNDGGYGHVAVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420
Db      361  ASAAAAGYRVGTPSPAGAVAVVNDGGYGHVAVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420
Qy      421  FNPGSVSYIYPN 432
Db      421  FNPGSVSYIYPN 432

RESULT 2
Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 99.2%; Score 2099; DB 2; Length 432;
Best Local Similarity 99.3%; Pred. No. 6.1e-88;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MKKRILSAVLNUGVTLSSATTLISAVKADDFDRAQLASQDSKINNLTAAQQAQAAQVNTIQ 60
Db      1 MKKRILSAVLNUGVTLSSATTLISAKADDFDRAQLASQDSKINNLTAAQQAQAAQVNTIQ 60
Qy      61 QVSALQTOQAEIQAENQRLEAQSATILGQOIQTLSKIVARNESLQQAARSQAQSNAAATSY 120
Db      61 QVSALQTOQAEIQAENQRLEAQSATILGQOIQTLSKIVARNESLQQAARSQAQSNAAATSY 120
Qy      121 INAINSKSVSDAINRVSAREVWSANEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180
Db      121 INAINSKSVSDAINRVSAREVWSANEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180
```

```
Qy      181  TTAQNTNALNTQQAQLEAAQLNLQAELETTAQQKATLVAKKAAAEAAARQAAAAQAAAEA 240
Db      181  TTAQNTNALNTQQAQLEAAQLNLQAELETTAQQKATLVAKKAAAEAAARQAAAAQAAAEA 240
Qy      241  KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
Db      241  KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
Qy      301  SQESTTATAAOPSPASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVNTWNGGOWA 360
Db      301  SQESTTATAAOPSPASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVNTWNGGOWA 360
Qy      361  ASAAAAGYRVGTPSPAGAVAVVNDGGYGHVAVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420
Db      361  ASAAAAGYRVGTPSPAGAVAVVNDGGYGHVAVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420
Qy      421  FNPGSVSYIYPN 432
Db      421  FNPGSVSYIYPN 432

RESULT 3
Q9AG98_STRMU
ID Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
AC Q9AG98;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
GN Name=saga;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21153617; PubMed=11254612;
RX DOI=10.1128/IAI.69.4.2493-2501.2001;
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
RT "Identification of stress-responsive genes in Streptococcus mutans by
differential display reverse transcription-PCR.";
RL Infect. Immun. 69:2493-2501(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RX DOI=10.1128/IAI.69.11.6987-6998.2001;
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
wall integrity and the maintenance of cell shape in Streptococcus
mutans.";
RL Infect. Immun. 69:6987-6998(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RA Chia J.S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
```



```

RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AF338445; AAK08104.1; -; Genomic DNA.
DR EMBL; AY046411; AAK94501.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5E47232E CRC64;

Query Match 98.0%; Score 2074.5; DB 2; Length 431;
Best Local Similarity 98.8%; Pred. No. 7.9e-87;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAAINTVAANO 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAAINTVAANO 180
QY 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQDKATLVAKAAAEAAEAQAAAAQAAEA 240
DB 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQDKATLVAKAAAEAAEAQAAAAQAAEA 240
QY 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSIYYPN 432
DB 421 FNPGSVSIYYPN 431

RESULT 4
Q938V3 STRMU
ID Q938V3 STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-SJ32;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=21481971; PubMed=11598068;
RX STRAIN-SJ32;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;

RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046410; AAK94500.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match 97.8%; Score 2070.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-86;
Matches 426; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAAINTVAANO 180
DB 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAAINTVAANO 180
QY 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQDKATLVAKAAAEAAEAQAAAAQAAEA 240
DB 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQDKATLVAKAAAEAAEAQAAAAQAAEA 240
QY 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSIYYPN 432
DB 421 FNPGSVSIYYPN 431

RESULT 5
Q8DWM3 STRMU
ID Q8DWM3 STRMU PRELIMINARY; PRT; 431 AA.
AC Q8DWM3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative secreted antigen GbpB/SagA; putative peptidoglycan
DE hydrolase.
GN Name=gbpB; OrderedLocusNames=SMU.22;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014855; RAN57811.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA695248CCD3E CRC64;

Query Match      97.8%; Score 2070.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-86;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAELETTAQQOKATLVAQKAAAEBAARQAAAAQA 240
Db 181 TTAQNTNALNTQQAOLEAAQLNLQAELETTAQQOKATLVAQKAAAEBAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAQAVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAQAVNN 300
QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
Db 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNOSIGNYRGW 420
Db 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNOSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
Db 421 FNPGSVSYIYPN 431

RESULT 6
Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_TaxID=1309;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8DB8C4609F CRC64;

Query Match      97.7%; Score 2066.5; DB 2; Length 431;
Best Local Similarity 98.4%; Pred. No. 1.8e-86;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAELETTAQQOKATLVAQKAAAEBAARQAAAAQA 240
Db 181 TTAQNTNALNTQQAOLEAAQLNLQAELETTAQQOKATLVAQKAAAEBAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAQAVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAQAVNN 300
QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
Db 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNOSIGNYRGW 420
Db 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNOSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
Db 421 FNPGSVSYIYPN 431

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE PcsB protein precursor.
DE PcsB protein precursor.
GN Name=PcsB;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_TaxID=1311;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183 (2001).
DR EMBL; AJ277292; CAC28144.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```



```

DR Sagalist; qbs0016; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.1%; Score 1229.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2e-48;
Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSATILSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 58
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSNATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 59 QVGALESQSELEAQAQLEAVSQQLGQEIQTLSKIVARNESLKQVRSQAQGN-LTNY 117
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 121 INAIINSKVSDAINRVSIAIREVVSANEMKMLHQBDKAAVEQKHQENQAANTVAANO 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180
QY 118 INTILNSKVSDAVNRVVAIREVVSANEMKMLAQEADKAALEAKQIENQNAINTVAA 177
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAQAAEA 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 240
QY 178 ALENKAALATQRAQLEAAQLESLAQITTVQNEKASLIQAQAEAAKAAEAQA 237
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAADNTQAAQTGSTDOSAAQAV 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 300
QY 238 KAAEAKAQAQESVAKAQA 285
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 285
QY 301 SDQ-----ESTTATA-----AQPSASASTAA-----VAANTSANTYPAG 336
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 336
QY 286 SSATTVAITTTATNEPKVTPSVTVKAVEAPKVVSVSTPRAVSKPVVRSVSSNTYP 345
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 345
QY 337 QCTWGVKSLAPWGVNGVNGGQWASAAAGVYRGTSPSAGAVVW--NDGGYGHVAVYT 394
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 394
QY 346 QCTWGAKSASWGVNGVNGWNGASARAAGVSGVTPRVGAVVAVPVDYGGYGHVAVYT 405
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 405
QY 395 GV-QGGQIQVQBANTYAGNQSIGNYRGWFPN---GSVSYIYPN 432
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 432
QY 406 SVANNSSIQVMSYAGNMSIGNYRGWFPNPSAGSVYIYPN 447
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 447

RESULT 10
Q5M6K4 STRT2
ID Q5M6K4_STRT2 PRELIMINARY; PRT; 474 AA.
AC Q5M6K4
DT 01-FEB-2005 (TRENBLrel. 29, Created)
DT 01-FEB-2005 (TRENBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Glucan binding protein (Pcgs).
GN Name=PcgsB; OrderedLocusNames=stcu0022;
OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=264199;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kalaukas S., Lapidus A., Goldsman E., Mazur M., Pusch G.D.,
RA Fontein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burtenu S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus."
RL Nat. Biotechnol. 22:1554-1558 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LMG 18311;
RT metagenome: complete genome sequence of a macrolide-resistant serotype

```

```

RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Leblond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus."
RL J. Bacteriol. 187:2737-2746 (2005).
DR EMBL; CP000023; AAV59752.1; -; Genomic DNA.
DR EMBL; AV730643; AAW82375.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BAC6A9E0A0A200D CRC64;

Query Match 51.4%; Score 1087.5; DB 2; Length 474;
Best Local Similarity 51.7%; Pred. No. 6.1e-42;
Matches 247; Conservative 72; Mismatches 108; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATILSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 57
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSNATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQARSQAQNNATSY 117
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 121 INAIINSKVSDAINRVSIAIREVVSANEMKMLHQBDKAAVEQKHQENQAANTVAANO 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 177
QY 118 INSTLNSKISSEAITRITAISKVVTANNLDLTQESDQKELAAQEEENQAANTTAANKS 177
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAQAAEA 236
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
QY 178 ELETTEAGLTQQAELEAAQVTLAELATAQNEKTSLSAKSTAESVAASTAASVAQSOA 237
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
QY 237 AAEAKA-----AAEAKALQEQAAQAAQAAANNNTQATDASDQAAADNTQAAQTG 287
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 287
QY 238 IAESEATAQVWASSEATSVASSEVAATSEVAQSPVSETS-TASEAAQEPASSETS 296
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 296
QY 288 DSTDQSAQAQVNN-----SDQESTTATAQPS-----ASSASTAAV-----AANT- 327
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 327
QY 297 EQPESAPAPVSEAPASVAPVATSEAPASVATSEAPASVATSEAPASVATSEAPASVAT 356
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 356
QY 328 -----SSANTYPAGCTWGVKSLAPWGVNGVNGGQWASAAAGVYRGTSPSAGAVVW 382
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 382
QY 357 KVSAASTPTNTPVGGCTWGVKSLAPWGVNGVNGWNGAKWIASAQAGHSVGTTPVAGIAVW 416
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 416
QY 383 -ND-GGYGHVAVYTGVOG-GQIQVQBANTYAGNQSIGNYRGWFPN-----GSVSYIYP 431
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 431
QY 417 PNDGGYGHVAVYTVTSAGANSIQVMSYAGNMSIGNYRGWFPDPTSSAHGGSVYIYP 474
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 474

RESULT 11
Q5XEL1 STRP6
ID Q5XEL1_STRP6 PRELIMINARY; PRT; 398 AA.
AC Q5XEL1
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OS Streptococcus pyogenes (serotype M6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbican K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., Deleo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus
RT metagenome: complete genome sequence of a macrolide-resistant serotype

```

```

DR EMBL: AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro: IPR007921; CHAP.
DR InterPro: IPR009148; SIDA.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS: PR01852; SIBAPROTEIN.
DR PROSITE: PS0911; CHAP; 1.
KW Complete prtome.
SQ
SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.4%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 5.4e-42;
Matches 233; Conservative 56; Mismatches 108; Indels 36; Gaps 6;

QY 1 MKKRLISAVLVSGVTILSSATTILSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRLISAVLVSGVTILGAAAT---VGAEDLSKIAKODSIISNLTTEQKAAQHQVSAQQA 57
QY 61 QVSALQTOQAELQAEORLEAOSATLGGQIQTLTSSKIVARNESLKQQAARSQAOKSNAATSY 120
DB 58 QVSSQLQSEQDKLTARTELEALSKEFEQIKALTSQIVARNEKLNKQARSAYKNNETSGY 117
QY 121 INAIINSKVSDAIRNRVSAREVVSANEKMLHQEQDKAAVEQKHQENQAINTVAANQE 180
DB 118 INALLNSKSIISDVNRLVAIRNAVSANAKLLEQOKADKVSLEBKQAANQTAINTIAANMA 177
QY 181 TTAQNTNALNTOQAELAAQLNLQABLTTAQOQKATLVQAQKAAAEAAQAAAAQAAEA 240
DB 178 MAEENQNTLRTQANLEAATAANLALQASATEDKANLVAQKEAAEKAAAEALAQEQAAKV 237
QY 241 KAAAEAKALQEQAAQAQAANNNNTQATDASDQAAAAADNTQAAOQTGSDTQSAQAQAVNN 300
DB 238 KA-----QEQAA-----QQAASVEAKSAIITPAQATPAQAASSNA 272
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGCQCTGWKVS LAPWGVNMGNGGQWA 360
DB 273 IFFAALTAPAA--PSARPPQI-----SYDSSNTYVPVCQCTGWAKSLAPWAGNMGNGGQWA 325
QY 361 ASAAAGAYRGVSTPSAGAVAVNWDGGYGHVAVYTGVO--GGQIOVQEAQVAGNOSIGNYRG 419
DB 326 YSAQAAGAYRTGTPTPMVGAIVAVNWDGGYGHVAVVVEVQSASSIRVWESNYSGRQYIADHRG 385
QY 420 WENPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 13
RPB3 STRP3
ID Q8P318_STRP3 PRELIMINARY; PRT; 398 AA.
AC Q8P318; Q7CFL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP80015, SpyM3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=SSI-1 / Serotype M3;
RC MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055 (2003).
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=MGAS315 / Serotype M3;

```

```

DR EMBL: AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro: IPR007921; CHAP.
DR InterPro: IPR009148; SIDA.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS: PR01852; SIBAPROTEIN.
DR PROSITE: PS0911; CHAP; 1.
KW Complete prtome.
SQ
SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.4%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 5.4e-42;
Matches 233; Conservative 56; Mismatches 108; Indels 36; Gaps 6;

QY 1 MKKRLISAVLVSGVTILSSATTILSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRLISAVLVSGVTILGTAATT---VGAEDLSKIAKODSIIISNLTTTEQKAAQHQVSAQQA 57
QY 61 QVSALQTOQAELQAEORLEAOSATLGGQIQTLTSSKIVARNESLKQQAARSQAOKSNAATSY 120
DB 58 QVSSLOSEQDKLTARTELEALSKEFEQEI KALTSQIVARNEKLNKQARSAYKNNETSGY 117
QY 121 INAIINSKSVDAIRNRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAINTVAANQE 180
DB 118 INALLNSKSI SDVNRLVAINRAVSANAKLLEQOKADKVSLEBKQAANQTAINTIAANMA 177
QY 181 TTAQNTNALNTOQAELAAQLNLQABLTTAQOKATLVQAQKAAAEAAQAAAAQAAAAEA 240
DB 178 MAEENQNTLRTQANLEAATAANLALQASATEDKANLVQAQKEAAEKAAAEALAQEQAAKV 237
QY 241 KAAAEAKALQEQAAQAQAQAANNNNTQATDASDQAAAAADNTQAAOQTGDSTDQSAQAQAVNN 300
DB 238 KA-----QEQAA-----QQAASVEAKSAIITPAQATPAQAQSSNA 272
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPACQCTGWKVS LAPWGVNMGNGGQWA 360
DB 273 IEPFALITAPAA-PSARPPQI-----SYDSSNTYVPVQCCTGWGAKSLAPWAGNMGNGGQWA 325
QY 361 ASAAAGAYRGVSTPSAGAVAVNWDGGYGHVAVYTGVO-GGQIOVQEAQVAGNOSIGNYRG 419
DB 326 YSAQAAGAYRTGTPTPMVGAIVAVNWDGGYGHVAVVVEQSSASIRVWESNYSGRQYIADHRG 385
QY 420 WENPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 13
RPB3 STRP3
ID Q8P318_STRP3 PRELIMINARY; PRT; 398 AA.
AC Q8P318; Q7CFL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP00015, SpyM3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RP [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RC MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055 (2003).
RP [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS315 / Serotype M3;

```

```

RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Berse S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL; BA000034; BAC63110.1; -; Genomic DNA.
DR EMBL; AB014136; AM78621.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.4%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 5.4e-42;
Matches 233; Conservative 56; Mismatches 108; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAVNTIQG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSTKIAKQDSIISNLTTEQKAAQNSALQA 57

QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAOKSNAATSY 120
DB 58 QVSSLSQEQDKLTARNTLEALSKRFEQIKALTSQIVARNESLKQARSAYKNNETSGY 117

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLHQODKAAVEQKHQENQAANTVAANO 180
DB 118 INALLNSKSIISDVNRLVAINRVANAKLEEQKADKVSLEEKQAQNTAINTAANNA 177

QY 181 TIAQNTNALNTQQAOLEAAQLNQLAELTTAQDQKATLVAQKAAAEFAAQAQAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKAAEAQAAQAAKV 237

QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTOSAAQAVNN 300
DB 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAAQSSNA 272

QY 301 SDOESTTATAOPSASSASTRAVAANTSSANTYPAGQCTGWKSLAPVGNWNGGOWA 360
DB 273 IEPALTAPAA-PSARPOT-----SYDESNTYPVGQCTWGAKSLAPWAGNNGGOWA 325

QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVYVTGVQ-GGQIQVQEANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGTPWGAIAVNDGGYGHVAVVVEVQSASSIRVMESYSGROYIADHRG 385

QY 420 WNPQSVSVIYPN 432
DB 386 WNPNTGVTFIYPH 398

RESULT 14
QSM212_STRT1
ID QSM212_STRT1 PRELIMINARY; PRT; 485 AA.
AC QSM212;
DT 01-FEB-2005 (TRENBLrel. 29, Created)
DT 01-FEB-2005 (TRENBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN Name=pcsb; OrderedLocusNames=str0022;
OS Streptococcus thermophilus (strain CNR9 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=299768;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,

RA Kulakauskas S., Lapidus A., Goltsman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui X., Masuy D., Hancy F., Burteau S., Boutry M., Delcour J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV61641.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 51.4%; Score 1087; DB 2; Length 485;
Best Local Similarity 50.9%; Pred. No. 6.6e-42;
Matches 249; Conservative 70; Mismatches 108; Indels 62; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYDSQIAATNNAISNLASQEQEAQAVATIOS 57

QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAOKSNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARODSLAKQARSQAQNTATSY 117

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLHQODKAAVEQKHQENQAANTVAANO 180
DB 118 INSLNSKSISEAITRTAISKVVTANNNDLLTKQESDQKELAAKQEEQAAINTTAAKS 177

QY 181 TIAQNTNALNTQQAOLEAAQLNQLAELTTAQDQKATLVAQKAAAEFAAQAQAAEA 236
DB 178 ELETTEAGLTQQAELEAAQVTLAAELATAQNEKTSLSVSAKSTABSVASTAAVAQSOA 237

QY 237 AAEAKA-----AAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTG 287
DB 238 IAESENTAQVVASSEATSVASSEVAATSEAVAQPSPTVSEIS-TASEAAQEPASSETS 296

QY 288 DSTQSAQAQAVNN-----SDQSTTATA-----AQPSSA----- 318
DB 297 EVQPESSAPVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPVSEAPVSAAPV 356

QY 319 -STRAVAANT-----SSANTYPAGQCTGWKSLAPVGNWNGGOWAASAAAGYRVG 371
DB 357 TSEAPAAETHKVSAASTPNTYPVGQCTGWKSLAPWAGNNGGOWAASAAAGYRVG 416

QY 372 STPSAGAVAVVNDGGYGHVAVYVTGVQ-GGQIQVQEANYAGNQSIGNYRGWFPN----- 423
DB 417 TTPVAGAIAPVNDGGYGHVAVYVTGSAGANSIQVMESYAGNMSISNYRGTFDPTSSAH 476

QY 424 -GSVSVIYP 431
DB 477 GGSVFIYP 485

RESULT 15
Q9A128_STRPY
ID Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; Q7BH59;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
```

RX	MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA	Ferretti J.J., McShan W.M., Ajdic D.N., Savic D.J., Savic G., Lyon K.,
RA	Primeaux C., Sezate S., Suvorov A.S., Kenton S., Lai H.S., Lin S.P.,
RA	Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RN	[2]
RP	NUCLEOTIDE SEQUENCE.
RA	Pagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT	"Identification and characterization of a novel secreted protein from
RT	group A streptococcus";
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AE006474; AAK331158.1; -- Genomic DNA.
DR	EMBL: AF319939; AAL73135.1; -- Genomic_DNA.
DR	InterPro; IPR007921; CHAP.
DR	InterPro; IPR009148; SibA.
DR	Pfam; PF05257; CHAP; 1.
DR	PRINTS; PR01852; SIBAPROTEIN.
DR	PROSITE; PS09111; CHAP; 1.
KW	Complete proteome; Signal.
FT	SIGNAL 1 23 Potential.
SQ	SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;
	Query Match 51.1%; Score 1081; DB 2; Length 398;
	Best Local Similarity 53.6%; Pred. No. le-48;
	Matches 232; Conservative 56; Mismatches 109; Indels 36; Gaps 6
Qy	1 MKKRLSAVLGVGTYLSATILSAVKADPDFDAQIASQSKINNLTAAQQAAQAQVNTIG 60
Dd	1 MKKRLSAVLGVGTYLGAATT---VGAEDELSTKIAKQDSIIISNLTFEQKAAQNQSALQA 57
Qy	61 QVSALQTQQAELQAEONBLEAQSATLGQOIOTLSSKIVARNESLKQARSASOKSNAATS 120
Dd	58 QVSSLQSEODKLTAENTLEALSKEPFEDEIKALTSGIARNEKLNQARSAYKNETSCTSY 117
Qy	121 INAIINSGSVSDAINRVSAIREVVSANEMKLHQEQFDKAAVEQKHENQAQINTVAANO 180
Dd	118 INALLNSKSIIDVVNRVLVAINRAVSANAKLLFEQKADKVSLFEKKQAANTATNTTAA 177
Qy	181 TIAONTVALNTCQAQLAAQNLQALHTAODOKATLVQAQAAAEEAARQAAAAQAAEA 240
Dd	178 MAENQNTLRTOQANLVAAVTANLALQLASATEDKANLVQAKEAEKAAREALAQAQAAKV 237
Qy	241 KAAAEAKALQEQAAQAAQAAAANNNTQATDASDQAAAAADNTQAAQTGDSTDOSAAQAVNN 300
Dd	238 KA-----QEQAA-----QQAASEAAKSATTPAQATPAQSSNA 272
Qy	301 SDQESTTATAAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPVGNWTWNGGGWA 360
Dd	273 IEPAALTAPAA--PSAGPQT-----SYOSSNTYPVGQCTWGAKSLAPAGNWNMGGGWA 325
Qy	361 ASAAAAAGYRVSGETSPAGAVAVMNDGGYGHVAVTVGVQ--GGIQVQEANYAGNQSIGNYRG 419
Dd	326 YSAQAAGYRGTGSTPMVGGAIVMNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYTADHRG 385
Qy	420 WFNPGSVSYIYPN 432
Dd	386 WFNPETGVTFIYPH 398

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 30.0232 Seconds
(without alignments)
1189.611 Million cell updates/sec

Title: US-10-797-821-32
Perfect score: 2116
Sequence: 1 MKGRILSAVLVSGVTLSAT.....SIGNVRGNFPGSVYIYPN 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/pCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	49.8	399	2	US-09-107-433-3230
2	1052	49.7	392	2	US-09-583-110-4374
3	679.5	32.1	461	1	US-08-186-222-2
4	581.5	27.5	525	2	US-09-107-532A-5095
5	555	26.2	210	2	US-09-222-938A-67
6	527	24.9	497	2	US-09-134-000C-5990
7	494.5	23.4	449	2	US-09-071-035-482
8	494.5	23.4	449	2	US-10-206-576-482
9	494.5	23.4	450	2	US-09-134-000C-5714
10	445.5	21.1	422	2	US-09-071-035-484
11	445.5	21.1	422	2	US-10-206-576-484
12	280.5	13.3	469	2	US-09-489-039A-13565
13	237	11.2	264	2	US-09-134-001C-5035
14	236	11.2	257	2	US-09-710-279-3244
15	236	11.2	267	2	US-09-134-001C-4539
16	230	10.9	477	2	US-09-902-540-11649
17	230	10.9	2310	2	US-08-874-923-120
18	224	10.6	1236	2	US-09-769-787-109
19	222.5	10.5	270	2	US-09-134-001C-5441
20	214.5	10.1	266	2	US-09-134-001C-5453
21	214	10.1	149	2	US-09-710-279-1682
22	214	10.1	157	2	US-09-710-279-2870
23	213	10.1	610	2	US-09-336-447A-11
24	213	10.1	610	2	US-09-952-267B-11
25	212.5	10.0	468	2	US-09-328-352-6321
26	211.5	10.0	639	2	US-09-902-540-14908
27	211.5	10.0	655	2	US-09-902-540-10005

28	211	10.0	440	2	US-08-302-756E-35	Sequence 35, Appl
29	206.5	9.8	1566	1	US-08-687-956A-23	Sequence 23, Appl
30	206	9.7	679	2	US-09-489-039A-12307	Sequence 12307, A
31	201.5	9.5	955	1	US-08-006-676B-1	Sequence 1, Appl1
32	201.5	9.5	955	1	US-08-282-845-2	Sequence 2, Appl1
33	201.5	9.5	955	1	US-08-428-414A-3	Sequence 3, Appl1
34	201.5	9.5	955	4	PCT-US94-00324-1	Sequence 1, Appl1
35	201	9.5	1070	2	US-09-902-540-13861	Sequence 13861, A
36	195	9.2	2107	2	US-09-949-016-7646	Sequence 7646, Ap
37	195	9.2	2107	2	US-09-949-016-7647	Sequence 7647, Ap
38	195	9.2	2115	2	US-09-296-662-33	Sequence 33, Appl
39	194.5	9.2	883	2	US-09-489-039A-12755	Sequence 12755, A
40	193.5	9.1	2101	1	US-08-466-390-4	Sequence 4, Appl1
41	193.5	9.1	2101	1	US-08-470-950-4	Sequence 4, Appl1
42	193.5	9.1	2101	1	US-08-467-781-4	Sequence 4, Appl1
43	193.5	9.1	2101	1	US-08-195-487-4	Sequence 4, Appl1
44	193.5	9.1	2101	1	US-08-483-924-4	Sequence 4, Appl1
45	193.5	9.1	2101	2	US-09-452-294-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1

US-09-107-433-3230
; Sequence 3230, Application US/09107433
; Patent No. 6800744

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3230:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...399

SEQUENCE DESCRIPTION: SEQ ID NO: 3230:


```
Best Local Similarity 35.6%; Pred. No. 1.9e-44; Mismatches 171; Conservative 104; Mismatches 137; Indels 69; Gaps 14;
1 MKKRILSVLVSGVTLTSAVKAADFDQAIAQSDSKINNLTAAQQAQAQVNTIQQ 60
Db 1 MKKKIISALMTSVILSAAPISGVVADT-NSDIAQDATISSAQSAQAQAQVDSLOS 59
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQNAATSY 120
Db 60 KVDLSQKQKQSTKAQIAKTESERKALNAQIATLINESIKERTKTLEAQAARSQVNSSATY 119
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAANTVAANQE 180
Db 120 MDVAVNSKSLTDVIOKVTAIATVSSANKOMLEQEQEKELSKQSTVTKKYNQFVLSQ 179
QY 181 TTAQNTNALNTQQAOLAEALQLLQAEELTTAQDQKATLVQAQAAAEAAQAAAAQA 240
Db 180 SLDQAQELTSQAELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQAKQAAYEA 239
QY 241 KAAAEKALQEAQAQAQAANNNNTQATDASDQQAADNTQAQDGTQDQ-----SAA 295
Db 240 Q-----QKEAAQAQAATAAKAVEAA-TSSASASSQAQPVSTSTDTNTTSNASAS 290
QY 296 QAVNSDQESTTATAAQPASASSASTAAVAANTSS-----ANTY 333
Db 291 NSSNSNSSSSSSSSSSSSSSSSNSNAGNTNSTGTGTGTTGGSGINSPIGNPY 350
QY 334 PAQOCT--WG-----VKSLAPVGVNGYGOWAASAAAAG--YRVGTPPSAGAVA 380
Db 351 AVGGCTDYVYQPAAGYIIRIMP-----GNGGQWASNGPAQGVHLVVGGAAP--GVTA 402
QY 381 V-----WMDGGYGHVAVYTVGQV-GQIQVQEAANYAGNQSIGNYRGNFPGSVSYIYP 431
Db 403 SSFSADFVGYANSPYGHVAIVKSVNSDGTITIKEGGY-GTTWGHGR-TVSASGVTFLMP 460
QY 432 N 432
Db 461 N 461

RESULT 4
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5095:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...525
; SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095
Query Match 27.5%; Score 581.5; DB 2; Length 525;
Best Local Similarity 30.4%; Pred. No. 8.1e-37;
Matches 160; Conservative 90; Mismatches 165; Indels 111; Gaps 14;
QY 1 MKKRILSVLVSGVTLTSAVKAADFDQAIAQSDSKINNLTAAQQAQAQVNTIQQ 60
Db 2 VKKSLISAVVWCMTLTAVASPIAAADDFDSQIQDQDKIADLNKQDADAQSDIDALES 61
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQNAATSY 120
Db 62 QVSEINTQAQDLAKQDTLRQESAQLVKDIADLQERIEKREDTIQKQAREAQVNTSSNY 121
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAANTVAANQE 180
Db 122 IDAVLNADSLADAIGRVQAMTTMVKANNDLMBEQQDKKAVEDKKAENDAKLKAENQA 181
QY 181 TTAQNTNALNTQQAOLAEALQLLQAEELTTAQDQKATLVQAQAAAEAAQAAAAQA 226
Db 182 ALESQKGLLSKQADLNVLKTSLAABQATAEKKADLNQKAAEAQAEQRIEQQLAEQ 241
QY 227 AARQAAAAQAQAAEAKAALQEAQAQAQAANNNNTQATDAS-----DQQAAD 279
Db 242 ARQAAQAEKAEAREQAEAEA--QATQASSTAQSSATEESSATQSSMTSESSSATQSS 298
QY 280 NTQAAQTGSDTQSAQAQVNSDQESTTA-----TAAQPSASSASTAAVAANTSS 329
Db 299 ATEESTTPESSTEESTAPESSESTTAPESSESTTAPESSESTTAPESSESTTAPESSTTE 358
QY 330 ANTYPA-----GQCTWGVKSLAP----- 347
Db 359 ESTTPATPTTSDQSDVTGNGT-GSSTPATPTTPEQPKVTPAPAPSGSVNGAAIVAE 417
QY 348 ---WYGN--YMGNG-----GQWASAAAAAGYRVG-STPSAGA 378
Db 418 AVKYIGTPVWGGKDPSCGFDSCGFYVYVMTGTRDIGGWTVPQESSAGTKISVQAKAGD 477
QY 379 VAVW-NDGGYGHVAVYTVGQGIQVQEAANYAGNQSIGNYRGNFNP 423
Db 478 LLFWGSQGGTTHVAIALG--GGQ-YIHAPQPGESVKVGSVQ-WFAP 519

RESULT 5
US-09-222-938A-67
; Sequence 67, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Chrisian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
```


Db 182 ELETQRDLSKQSEINVMKASLALAEQSSAESKAGLEKQKAAAEQARLAAEQKAAE 241
 QY 227 AARQAAAAQAAAEAKAAAEAKALQEQAAQAAANNNNTQATDASDQAAA-----AADNT 281
 Db 242 KAKQAAAKPAKAEVK--AEAPVASSSTTEAQAPASSSS--ATESSTQQTTPSTPSTDNS 297
 QY 282 QAAQCGDSTDSAAQAV-----NNSDQESTTATAAQPSASSASTAAVAANTSSANTYPA 335
 Db 298 ATENTGSSSSSEPOVQFTPTSDNGNGGQGTGTTPTPTPAPSADPTINALNVLROSL 357
 QY 336 GCGTWGKSLAPWNGVNGWGQOWAASAAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTG 395
 Db 358 G-----LRPVV---WDAGLAASATARAQAEAGGIPNDH-----WSRG---DEVIAIMW 400
 QY 396 VQGGQIQV---QEANYAGNQSIGNRYGM-FNPG 424
 Db 401 APGNSVINAWYNETNMVTASGSG-HRDWEINPG 432

RESULT 10
 US-09-071-035-484
 ; Sequence 484, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,035
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: A. Anders Brookes
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB369P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 484:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 422 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-071-035-484

Query Match 21.1%; Score 445.5; DB 2; Length 422;
 Best Local Similarity 29.8%; Pred. No. 1.9e-26;
 Matches 127; Conservative 80; Mismatches 168; Indels 51; Gaps 12;
 QY 28 DFDQAIASQDSKINLTAQQAQAAQVNTIQGVYALQTAQAEQAEORLEAQASATLG 87
 Db 1 DNVDRKKEIKQKESLAKQGDLSAQVSSLEAEVSSVFDSEALREKQTLKASEQLQ 60
 QY 88 QQIOTLSSKIVARNESLKKQQAASQKSNAAATSYINAINSKSVSDAINRVSAIREVWSAN 147

Db 61 QBITNLNQRIEKRNEAIKNQARDVQVNGQSTTMLDVAILDADSVADAISRQVQAVSTIVSAN 120
 QY 148 EKMLHQRODQKAAAYEOKHQENQAAINTVAANQETIAQNTNALNTQQAQLEAAQLNLAEL 207
 Db 121 NDLMOQOKEDQAVVDDKKAENEKVKQLQEALETQRDLSKQSEINVMKASLALAEQ 180
 QY 208 TTAODOKATLVAQKAAAE-----EAAQAAAAQAAAAEAKAAAEAKALQEOA 253
 Db 181 SSAESSKAGLEKQKAAAEAEQARLAAEQKAAAEKAKQAAKPAKAEVK--AEAPVASSST 238
 QY 254 AQAQAAANNNNTQATDASDQAAA-----AADNTQAAQTGSDTDSQAAAQAV-----NNSD 302
 Db 239 TEAQAPASSSS--ATESSTQQTTPSTPSTDNSATENTGSSSSSEPOVQFTPTSDNGNGG 296
 QY 303 QESTTATAAQPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWNGVNGWGQOWAAS 362
 Db 297 QTGGGTVTPTPTPAPSADPTINALNVLROSLG-----LRPVV---WDAGLAASAT 345
 QY 363 AAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTGVOGGIQV---QEANYAGNQSIGNRYGM 419
 Db 346 ARAAQVEAGGIPNDH-----WSRG---DEVIAIMWAFGNSVINAWYNETNMVTASGSG-HRD 398
 QY 420 W-FNPG 424
 Db 399 WEINPG 404

RESULT 11
 US-10-206-576-484
 ; Sequence 484, Application US/10206576
 ; Patent No. 6913907
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et al.
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 497
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-R
 ; COMPUTER: Dell Latitude
 ; OPERATING SYSTEM: Windows 98
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/206,576
 ; FILING DATE: 29-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/071,035
 ; FILING DATE: 1998-05-04
 ; APPLICATION NUMBER: US 60/046,655
 ; FILING DATE: 1997-05-16
 ; APPLICATION NUMBER: US 60/044,031
 ; FILING DATE: 1997-05-06
 ; APPLICATION NUMBER: US 60/066,009
 ; FILING DATE: 1997-11-14
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hyman, Mark J.
 ; REGISTRATION NUMBER: 46,789
 ; REFERENCE/DOCKET NUMBER: PB369PID1
 ; INFORMATION FOR SEQ ID NO: 484:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 422 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 484:
 ; US-10-206-576-484

Query Match 21.1%; Score 445.5; DB 2; Length 422;
Best Local Similarity 29.8%; Pred. No. 1.9e-26;
Matches 127; Conservative 80; Mismatches 168; Indels 51; Gaps 12;

QY 28 DFDIAQIASODSKINLTAQQAQAQAVNTIGQVSAQTQQAELQAEQNRLEAQSATIG 87
DB 1 DNVDKIEKNOEISLAKAQGDLSQVSSLEAEVSSVFDESMALREQQTLLKAKSEQIQ 60
QY 88 QOIQTLSSKIVARNESLQQAARSQAQSNAAATVINAINSKVSDAINRVSAIREVVSAN 147
DB 61 QEITWLNQRIEKNRAIKNQARDVQVNGOSTTMDLDAVDAISRVQAVSTIVSAN 120
QY 148 EKMLHQEQBDKAABEQKHQENQAINTVAANQETIAQNTNALNTQQAQLAQAQLNLOAEL 207
DB 121 NDLMQOQKEDQAVVDKKAENEKKVKQLEATEAELETKRQDLLSKQSELNVKMSLAEQ 180
QY 208 TTAQOKATLVAKAAAE-----EAAQAAQAQAQAQAQAQAQAQAQAQAQAQAQA 253
DB 181 SSAESSKAGLEKQKAAAEQAARLAAEQKAAAEKAKQAQAQAQAQAQAQAQAQAQA 238
QY 254 AQAQAANNNNTQATDASDQAAA-----AADNTQAAQTGSDTQSAQAQAV-----NNSD 302
DB 239 TEAQAPASSS--ATESSTQQTETTTPTSDNSATENTGSSSEQPVQPTTSPDNGNGG 296
QY 303 QESTTATAAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGWGGQWAAS 362
DB 297 QTGGGTVTPTPTPAPSADPTINALNVLRSGLG-----LRPVV---WDAGLAASAT 345
QY 363 AAAAGYRGVSTPSAGAVAVNDGGGHVAVYTVGVGGQIQV---QEAQVAGNQSIGNYRG 419
DB 346 ARAAQVEAGGIENDH-----WSRG--DEVIAIMWAFGNSVIMAWYNETNMVTASGSG-HRD 398
QY 420 W-FNPG 424
DB 399 WEINEP 404

RESULT 12
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

Query Match 13.3%; Score 280.5; DB 2; Length 469;
Best Local Similarity 28.9%; Pred. No. 1.2e-13;
Matches 113; Conservative 56; Mismatches 165; Indels 57; Gaps 12;

QY 31 DAQIASQDSKINLTAQQAQ-----AAQAVNTTIGQVSAQTQQAELQAEQNR 78
DB 84 DAIMVDPGAVNNVNRQQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 143
QY 79 LEAQSATLGGQQTQLTSSKIVARNESLQQAARSQAQSNAAATVINAINSKVSDAINRV 138
DB 144 LQAQEAQAEKQ-----QKQEAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 190
QY 139 AIREVVSAERKMLHQEQBDKAABEQKHQENQAINTVA-ANQETIAQNTNALNTQQAQL 197
DB 191 AEAKAKAQA 250

QY 198 AA---QLNLOAELTTAQDQKATLVAKAAAEBAARQAAAAQAQAQAQAQAQAQAQAQA 254
DB 251 AAAAKKAQQAQAEKKAQA 309
QY 255 QQAQAANNNNTQATDASDQAAAADNTQAAQTGSDT-----QSAQAQAVNNSDQES 305
DB 310 AAEKAAAD---KAAKAAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 366
QY 306 TTATAAQPSSASTA-----AAVAANTSSANTYPAGQCTGWGKSLAP--WVGNY 352
DB 367 AAAAGSGMTKKSASGADINNVAGQIKSAIESKFYDASSYAGTCTLRK-LAPDGLLLNI 425
QY 353 WNGCGWA-ASAAAAAGYRGVSTPSAGAVAVW 382
DB 426 QSEGGDPALCQAALAAARQAQKFPKPPSQAVY 456

RESULT 13
US-09-134-001C-5035
; Sequence 5035, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5035
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match 11.2%; Score 237; DB 2; Length 264;
Best Local Similarity 34.7%; Pred. No. 1.3e-10;
Matches 60; Conservative 27; Mismatches 70; Indels 16; Gaps 5;

QY 260 ANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNNSDQESTTATAAQPSSASS 319
DB 104 SNYNNYQ-----SNNTQSQTPTQPTGGLGASYSTSSNVHVTTSA-PSSNGVS 151
QY 320 TAAVAANTSSANTYPAGQCTGWV-KSLAPWGVNNGWGGQWAASAAAAAGYRGVSTPSAGA 378
DB 152 LS--NARSAGNLYTSGQCTYVYFDRVGKIGSTWGNANNWANAARSGYTVNNSPAKGA 209
QY 379 VAVWNDGGYGHVAVYTVG-QGQIQVQEAQVAGNQSIGNYRGWFPNGVSVYIY 430
DB 210 ILQTSQGAIGHVAVYVEGVNSNGSIRKSENNYGHGAGVVTSTRTISASQAASNY 262

RESULT 14
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 92.071 Seconds
(without alignments)
1960.467 Million cell updates/sec

Title: US-10-797-821-32

Perfect score: 2116

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRWGWFNPGSVSIYPN 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA_Main:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
 - 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
 - 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
 - 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2116	100.0	432	US-10-383-930-32	Sequence 32, Appl
2	2116	100.0	432	US-10-797-821-32	Sequence 32, Appl
3	2099	99.2	432	US-10-383-930-31	Sequence 31, Appl
4	2099	99.2	432	US-10-797-821-31	Sequence 31, Appl
5	2074.5	98.0	431	US-10-383-930-30	Sequence 30, Appl
6	2074.5	98.0	431	US-10-797-821-30	Sequence 30, Appl
7	2070.5	97.8	431	US-10-383-930-29	Sequence 29, Appl
8	2070.5	97.8	431	US-10-797-821-29	Sequence 29, Appl
9	2066.5	97.7	431	US-10-383-930-33	Sequence 33, Appl
10	2066.5	97.7	431	US-10-797-821-33	Sequence 33, Appl
11	1081	51.1	398	US-10-474-792-600	Sequence 600, App
12	1053	49.8	392	US-10-472-928-4652	Sequence 4652, Ap
13	1053	49.8	399	US-10-617-320-3230	Sequence 3230, Ap
14	585.5	27.7	524	US-10-617-320-3230	Sequence 3230, Ap
15	555	26.2	210	US-10-154-251-67	Sequence 67, Appl
16	494.5	23.4	449	US-09-071-035-482	Sequence 482, App
17	494.5	23.4	449	US-10-206-576-482	Sequence 482, App
18	494.5	23.4	449	US-10-912-362-482	Sequence 482, App
19	445.5	21.1	422	US-09-071-035-484	Sequence 484, App
20	445.5	21.1	422	US-10-206-576-484	Sequence 484, App
21	445.5	21.1	422	US-10-912-362-484	Sequence 484, App
22	320	15.1	630	US-10-494-674-6	Sequence 6, Appli
23	314	14.8	609	US-09-738-626-5197	Sequence 5197, Ap
24	300.5	14.2	609	US-10-501-282-3184	Sequence 3184, Ap
25	293.5	13.9	440	US-10-501-282-3056	Sequence 3056, Ap
26	291	13.8	422	US-10-501-282-3054	Sequence 3054, Ap
27	282	13.3	377	US-10-501-282-3052	Sequence 3052, Ap

ALIGNMENTS

RESULT 1

US-10-383-930-32

; Sequence 32, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 32

; LENGTH: 432

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-10-383-930-32

Query Match 100.0%; Score 2116; DB 4; Length 432;

Best Local Similarity 100.0%; Pred.No.1.9e-121; Mismatches 0; Indels 0; Gaps 0;

Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

Db 1 MKKRILSAVLVSGVTLSATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENORLEAQSAITLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120

Db 61 QVSALQTOQAELOAENORLEAQSAITLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120

QY 121 INAIINSSVSDAINRVSAIREVVSANESKMLHQQDDKAAVEQKHQENQAALNTVAANOE 180

Db 121 INAIINSSVSDAINRVSAIREVVSANESKMLHQQDDKAAVEQKHQENQAALNTVAANOE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEAAARQAQAAQA 240

Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEAAARQAQAAQA 240

QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAQGTGSDTDSAAQAVNN 300

Db 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAQGTGSDTDSAAQAVNN 300

QY 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGOWA 360

Db 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGOWA 360

```
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420

QY 421 FNPGSVSIYPN 432
Db 421 FNPGSVSIYPN 432

RESULT 2
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32
```

```
Query Match 100.0%; Score 2116; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.9e-121;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALOTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
Db 61 QVSALOTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAAANQE 180

QY 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 240
Db 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300

QY 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
Db 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420
```

```
QY 421 FNPGSVSIYPN 432
Db 421 FNPGSVSIYPN 432

RESULT 3
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31
```

```
Query Match 99.2%; Score 2099; DB 4; Length 432;
Best Local Similarity 99.3%; Pred. No. 2.1e-120;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALOTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
Db 61 QVSALOTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAAANQE 180

QY 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 240
Db 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300

QY 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
Db 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQAEYAGNOSIGNYRGW 420

QY 421 FNPGSVSIYPN 432
Db 421 FNPGSVSIYPN 432
```

```
RESULT 4
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
```

FILE REFERENCE: 25669-020
CURRENT APPLICATION NUMBER: US/10/797,821
CURRENT FILING DATE: 2004-03-09
PRIOR APPLICATION NUMBER: 10/383,930
PRIOR FILING DATE: 2003-03-07
PRIOR APPLICATION NUMBER: 60/363,209
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: 60/402,483
PRIOR FILING DATE: 2002-08-08
PRIOR APPLICATION NUMBER: 09/290,049
PRIOR FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
PRIOR FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31
LENGTH: 432
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match 99.2%; Score 2099; DB 5; Length 432;
Best Local Similarity 99.3%; Pred. No. 2.1e-120;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIG 60
DB 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIG 60
QY 61 QVSALQTOQAEIQAENQRLQAQSAITLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB 61 QVSALQTOQAEIQAENQRLQAQSAITLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120
QY 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQEQDQKAAVEQKQENQAAINTVAANOE 180
DB 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQEQDQKAAVEQKQENQAAINTVAANOE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAKAAAEAEARQAAAAQAAEA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGQWA 360
DB 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGQWA 360
QY 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSIYYPN 432
DB 421 FNPGSVSIYYPN 432

RESULT 5
US-10-383-930-30
Sequence 30, Application US/10383930
Publication No. US20040127400A1
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J
APPLICANT: Taubman, Martin A
TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
FILE REFERENCE: 25669-018
CURRENT APPLICATION NUMBER: US/10/383,930
CURRENT FILING DATE: 2003-03-07
PRIOR APPLICATION NUMBER: 60/402,483
PRIOR FILING DATE: 2002-08-08

PRIOR APPLICATION NUMBER: 60/363,209
PRIOR FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.2
SEQ ID NO 30
LENGTH: 431
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-383-930-30
Query Match 98.0%; Score 2074.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 6.6e-119;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIG 60
DB 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIG 60
QY 61 QVSALQTOQAEIQAENQRLQAQSAITLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB 61 QVSALQTOQAEIQAENQRLQAQSAITLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120
QY 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQEQDQKAAVEQKQENQAAINTVAANOE 180
DB 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQEQDQKAAVEQKQENQAAINTVAANOE 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAKAAAEAEARQAAAAQAAEA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 299
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGQWA 360
DB 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRGW 419
QY 421 FNPGSVSIYYPN 432
DB 421 FNPGSVSIYYPN 431
RESULT 6
US-10-797-821-30
Sequence 30, Application US/10797821
Publication No. US20050031633A1
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J
APPLICANT: Taubman, Martin A
TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
FILE REFERENCE: 25669-020
CURRENT APPLICATION NUMBER: US/10/797,821
CURRENT FILING DATE: 2004-03-09
PRIOR APPLICATION NUMBER: 10/383,930
PRIOR FILING DATE: 2003-03-07
PRIOR APPLICATION NUMBER: 60/363,209
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: 60/402,483
PRIOR FILING DATE: 2002-08-08
PRIOR APPLICATION NUMBER: 09/290,049
PRIOR FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
PRIOR FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 30

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match      98.0%; Score 2074.5; DB 5; Length 431;
Best Local Similarity 98.8%; Pred. No. 6.6e-119;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60
DB 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60

QY 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120
DB 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240

QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300

QY 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
DB 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420

QY 421 FNPGSVSVIYPN 431
DB 420 FNPGSVSVIYPN 431

RESULT 7
US-10-930-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-930-930-29

Query Match      97.8%; Score 2070.5; DB 4; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-118;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60
DB 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60

QY 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120
DB 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240

QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300

QY 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
DB 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420

QY 421 FNPGSVSVIYPN 431
DB 420 FNPGSVSVIYPN 431
```

```
RESULT 8
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match      97.8%; Score 2070.5; DB 5; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-118;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60
DB 1 MKKRILSAVLVSGVTLSSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60

QY 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120
DB 61 QVSALOTQOAELOAENORLEAQSATLGOOIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAAEA 240

QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTOSAAQAQVNN 300

QY 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
DB 301 SDQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNQSIGNYRGW 420

QY 421 FNPGSVSVIYPN 431
DB 420 FNPGSVSVIYPN 431
```

```
Db 121 INAIINSKSVDAINRVSAIREVVSANEKXLOQOEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
Db 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
QY 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 300
Db 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 299
QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
Db 300 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 419
QY 421 FNPGSVSYIYPN 432
Db 420 FNPGSVSYIYPN 431

RESULT 9
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33

Query Match 97.7%; Score 2066.5; DB 4; Length 431;
Best Local Similarity 98.4%; Pred. No. 2e-118;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQQ 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQQ 60
QY 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEKXLMHQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKSVDAINRVSAIREVVSANEKXLMHQEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
Db 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
QY 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 300
Db 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 299
QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
Db 300 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 420
Db 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
```

```
Db 300 SDOESTTATAOPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 419
QY 421 FNPGSVSYIYPN 432
Db 420 FNPGSVSYIYPN 431

RESULT 10
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match 97.7%; Score 2066.5; DB 5; Length 431;
Best Local Similarity 98.4%; Pred. No. 2e-118;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQQ 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQQ 60
QY 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120
QY 121 INAIINSKSVDAINRVSAIREVVSANEKXLMHQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKSVDAINRVSAIREVVSANEKXLMHQEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
Db 181 TTAQNTNALNTQOAELEAAQLNLQAELETTAQOKATLVAKAAAEAEARQAAQAQAAEA 240
QY 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 300
Db 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQAVNN 299
QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
Db 300 SDOESTTATAOPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 420
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVGGGQIQVQEANYAGNOSIGNYRGW 419
```

```
; 421 FNPQSVSYIYPN 432
; |||||
Db 420 FNPQSVSYIYPN 431

RESULT 11
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 600
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-600

Query Match 51.1%; Score 1081; DB 5; Length 398;
Best Local Similarity 53.6%; Pred. No. 3.2e-58;
Matches 232; Conservative 56; Mismatches 109; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTTQ 60
DB 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSTKIAKQDSIIISNLTTEQAAQNVSAQ 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQSNAAATSY 120
DB 58 QVSSLOSQEQDKLTARNTLEALSKEPEQEIKAITSQIVARNEKLNQARSAYKNNETSGY 117

QY 121 INAINSKSVSDAINRVSAIREVVSANEMKMLHQEQDQKAAVEQKHQHOENQAINTVAANQE 180
DB 118 INALLNSISIDVNVRLVAINRAVSANAKLLFEQKADKVSLEEKQAANTQTANTITAAANA 177

QY 181 TIAQNTNALNTQOALEAQNLOAELTTAODOKATLVAQKAAAEAEARQAAQAAAE 240
DB 178 MABENQNTLRTQOANLVATANTLALQLASATEDKANLVAQKEAEKAAAEALAQEQAAKV 237

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAAADNTQAAQTGDSQDQSAQAQVNN 300
DB 238 KA-----QEQA-----QQAASVEAAKSAITPAPQATPAAQSSNA 272

QY 301 SDQESTTTATAOPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
DB 273 IEPAAALTAPAA-PSAGPQT-----SYDSSNTYPVPGQCTGWGAKSLAPWAGNNGGQWA 325

QY 361 ASAAAAGYRVGSTRSAGAVAVNWDGGYGHVAVVTGVO-GGQIQVOEANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGTPVMVGAIAVWNDGGYGHVAVVVEVQSASSIRVMESNYSGRYTADHRG 385

QY 420 WFNPGSVSYIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 12
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
```

```
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.B+01)
US-10-472-928-4652

Query Match 49.8%; Score 1053; DB 5; Length 392;
Best Local Similarity 49.8%; Pred. NO. 1.6e-56;
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTTQ 60
DB 1 MKKKILASLLSLSTVMVSVQAVLTTTAHAETTDKIAAQDNKISNLTAAQQQEAQKQVDQIQE 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQSNAAATSY 120
DB 61 QVSAIQAEQSNLOAENDRLQAESKLEGEITELSKNIVSRNQSLEKQARSAGTNGAVTSY 120

QY 121 INAINSKSVSDAINRVSAIREVVSANEMKMLHQEQDQKAAVEQKHQHOENQAINTVAANQE 180
DB 121 INTIVNSKSITEAISRAVAMSEIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ 180

QY 181 TIAQNTNALNTQOALEAQNLOAELTTAODOKATLVAQKAAAEAEARQAAQAAAE 240
DB 181 KLADDAQALTTQOAELEKAAELSLAAEKATAGEKASLLEQKAAAEAEARAAVAEAAAYKE 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAAADNTQAAQTGDSQDQSAQAQVNN 300
DB 241 KRASQQQSV-----LASANTNLTAQOVAVSESAAPVRAKVRPT----- 279

QY 301 SDQESTTTATAOPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360
DB 280 -----YSTNASSYPITGECTGWGVKTLAPWAGDYWNGGAQWA 314

QY 361 ASAAAAGYRVGSTRSAGAVAVNWDGGYGHVAVVTGVOG-GQIQVOEANYAGNQSIGNYRG 419
DB 315 TSAAAGPRTGSTPOVGAIACWNDGGYGHVAVVTVAVESTTTRIQVSESNYAGNRTIGNHRG 374

QY 420 WFNP-----GSVSYIY 430
DB 375 WFNPTTSEGFTVIY 390

RESULT 13
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
```

OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
FILING DATE: 10-Jun-1998
FILING DATE: May 12, 1998
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3230:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...399
SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

Query Match 49.8%; Score 1053; DB 5; Length 399;
Best Local Similarity 49.8%; Pred. No. 1.7e-56; Indels 52; Gaps 4;
Matches 217; Conservative 74; Mismatches 93;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQVNTTQ 60
DB 8 MKKILASLLSTVMVSVQVAVLTAAHTTDDKIAAQDNKISNLTAAQQAQVNDQIQE 67
QY 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNATSY 120
DB 68 QVSAIQAEQSNLQAEADRLOAESKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 127
QY 121 INALINSKVSDAINRVSAIREVVSANEMKLMHQODKAAVEQKHQENQAQNTVAANQE 180
DB 128 INTIVNSKSIITAIRVAAAMSVIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ 187
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAOPQKATLVAKAAAEAEARQAAAAQAAEA 240
DB 188 KLADDAQALTTQAEKKAELSIAEKATAEKESALLEQKAAAEARAAVAEAAEAYKE 247
QY 241 KAAAEKALQEQAAQAAANNNTQATDASDQQAADNTQAAQGTGSDTQSAQAQAVNN 300
DB 248 KRAAQOQSV-----LAGANTLTAQVQAVESAAAPVRAKVRPT----- 286
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGOCITWGVKSLAPWVNTYWGNGGOWA 360
DB 287 -----YSTNASSYPIGECTWGVKTLAPWAGDYWGNGAOWA 321
QY 361 ASAAAAGVVGSTPSAGAVVNDGGYGHVAVVTGVQV-GQIQVEANYAGNOSIGNYRG 419
DB 322 TGAAGAAGFTGTPQVGAICNDGGYGHVAVVTAVESTTRIQVESNYAGNRTIGNHNG 381
QY 420 WFNPE-----GSVSYTY 430
DB 382 WFNPTTTSSEGFVTYIY 397

RESULT 14
US-10-282-122A-57658

Sequence 57658, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57658
LENGTH: 524
TYPE: PRT
ORGANISM: Enterococcus faecium
US-10-282-122A-57658

Query Match 27.7%; Score 585.5; DB 4; Length 524;
Best Local Similarity 30.6%; Pred. No. 8.6e-28;
Matches 161; Conservative 89; Mismatches 165; Indels 111; Gaps 14;
QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQVNTTQ 60
DB 1 MKKSLISAVMCSWTLTAVASPIAAADDFDSIQQQDKIADLNKQQAQSDALES 60
QY 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNATSY 120
DB 61 QVSEINTQADLLAKQDTRQESAQLVKDIADLQERIEKREDTIQKQAREQVNTSSNY 120
QY 121 INALINSKVSDAINRVSAIREVVSANEMKLMHQODKAAVEQKHQENQAQNTVAANQE 180
DB 121 IDAVLNADSLADAIGRVQAMTTVMKANNNDLMEQQKQDKKAVEDKKAENDAKLKEAENQA 180
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAODOKATLVAKA-----AAEE 226
DB 181 ALESQKGLLSKQADLNVLKTSLAEQATAEKDKADLNKQKAAEAEQARIREQORLADQ 240
QY 227 AARQAAAAQAAAEKAAAEKALQEQAAQAAANNNTNTQATDAS-----DQAAAAAD 279
DB 241 ARQQAQAEKAEKAEQAEAEAE---QATQASSTAOSSATESSATQSSMTTESSATQSS 297
QY 280 NTQAAQTGDSTQSAQAQAVNNSDQESTTA-----TAAQPSASSASTAAVAANTSS 329


```
Db      298 ATEESTPESSTESTAPESSTESTAPESSTESTTVPESSTESTTVPESSTTE 357
Qy      330 ANTPA-----GCTWGVKSLAP-----
Db      358 ESTTPAPTPTDQSVDTGNGT-GSSTPAPTPTPEQPKPVPAPAPSGSVNGAAIVAE 416
Qy      348 ---WGN--YWGNG-----GWAASAAAAGYRVG-STPSAGA 378
Db      417 AYKYGTPYVWVGKDPGSGFSGFTRYVYVMQVTGRDIGGWTVPQBSAGTKISVQAKAGD 476
Qy      379 VAVW-NDGGYGHVAVYVVGQGOIOVOEANYAGNOSIGNYRGWFNP 423
Db      477 LWFWSQGGTYHVAIALG--GGQ-YIHAPQGESVKGVSQ-WFAP 518
```

RESULT 15

```
US-10-154-251-67
; Sequence 67, Application US/10154251
; Publication No. US20030092024A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Pritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 06286-060002
; CURRENT APPLICATION NUMBER: US/10/154,251
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 10/154,251
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-154-251-67
```

```
Query Match      26.2%; Score 555; DB 4; Length 210;
Best Local Similarity 54.8%; Pred. No. 2.2e-26;
Matches 115; Conservative 44; Mismatches 51; Indels 0; Gaps 0;
```

```
Qy      1 MKKRILSAVLVSGVTILSSATTLISAVKADDPDAQLASODSKINNLTAAQQAAQAVNTIQG 60
Db      1 MKKKILASLLSTVMVSVQAVLTTAHAETDDKIAAQDNKISNLTAAQQAAQAVNTIQG 60
Qy      61 QVSALQTOQAELEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQSNAAATSY 120
Db      61 QVSAIQAEQSNLQAEENDRLQAESKLEGEITELSKNIVSRNQSLEKQARSQAQSNAAATSY 120
Qy      121 INAIINSKSVDAINRVSAIRVVSANEKMLHQEQODKAAVEQKHQENQAANTVAANQE 180
Db      121 INTIVNSKSITEAISRVAAMSIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQQ 180
Qy      181 TTAQNTNALNTQQAQLAEALNLAELTTA 210
Db      181 KLADDAQALTTKQAELEKAAELSLAAEKATS 210
```

Search completed: February 15, 2006, 18:34:53
Job time : 93.071 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2070.5	97.8	431	7	US-11-052-554A-210	Sequence 210, App
2	1087	51.4	398	7	US-11-052-554A-352	Sequence 252, App
3	436.5	20.6	211	7	US-11-052-554A-352	Sequence 352, App
4	410	19.4	544	7	US-11-052-554A-358	Sequence 358, App
5	236	11.2	257	6	US-10-793-626-3244	Sequence 3244, App
6	228	10.8	971	7	US-11-052-554A-3	Sequence 3, Appl
7	226.5	10.7	394	7	US-11-052-554A-79	Sequence 79, Appl
8	224	10.6	1236	6	US-10-873-528-109	Sequence 109, App
9	222.5	10.5	270	6	US-10-485-517-413	Sequence 413, App
10	214	10.1	149	6	US-10-793-626-1682	Sequence 1682, Ap
11	214	10.1	157	6	US-10-793-626-2870	Sequence 2870, Ap
12	193.5	9.1	2101	6	US-10-857-780-23	Sequence 23, Appl
13	188.5	8.9	1562	7	US-11-052-554A-211	Sequence 211, App
14	184.5	8.7	1095	6	US-10-793-626-3154	Sequence 3154, Ap
15	183.5	8.7	758	6	US-10-485-517-144	Sequence 144, App
16	183.5	8.7	1448	6	US-10-485-517-212	Sequence 212, App
17	180.5	8.5	1122	6	US-10-467-657-6112	Sequence 6112, Ap
18	180.5	8.5	1586	6	US-10-821-234-901	Sequence 901, App
19	178.5	8.4	1410	6	US-10-878-556A-136	Sequence 136, App
20	178	8.4	330	6	US-10-485-517-415	Sequence 415, App
21	177.5	8.4	1565	6	US-10-467-657-2704	Sequence 2704, Ap
22	177	8.4	1107	6	US-10-485-517-145	Sequence 145, App
23	176	8.3	1290	6	US-10-485-517-141	Sequence 141, App
24	175	8.3	1126	6	US-10-485-517-248	Sequence 248, App
25	174	8.2	1634	7	US-11-019-711-49	Sequence 49, Appl

Qy	60	QGVSA	QTOQA	ELQ----	AENOR	LEAQS-----	ATL	86
		:	:	:	:	:	:	
Db	62	VAIWSE	ENGQD	DLKWYH	ASNDGS	NQLTVHFA	ENHGS	121
		:	:	:	:	:	:	
Qy	87	GOQLOT	LS	SSKI	VARNES	LKQ--	-QARS	139
		:	:	:	:	:	:	
Db	122	GKRKL	SLS----	APQLS	LKQ	QGLQLF	SKL	177
		:	:	:	:	:	:	
Qy	140	-----	-----	-----	-----	-----	-----	178
		:	:	:	:	:	:	
Db	178	NTLAGY	ANHKG	YGT	VHVHTY	LKQ	SKMIP	237
		:	:	:	:	:	:	
Qy	179	-----	-----	-----	-----	-----	-----	206
		:	:	:	:	:	:	
Db	238	VPPYIS	VAIP	PVWSE	QGD	LKWYQ	ATKVAD	297
		:	:	:	:	:	:	
Qy	207	-----	-----	-----	-----	-----	-----	249
		:	:	:	:	:	:	
Db	298	LSKKL	DGLG	ETHF	NVP	SI	INVED	348
		:	:	:	:	:	:	
Qy	250	QEOAAQ	AAAA	NNNT	-----	QATD	ASDQ	293
		:	:	:	:	:	:	
Db	349	QSTIS	AAVMS	DANQ	ANLYE	AKQLANG	KAAIT	408
		:	:	:	:	:	:	
Qy	294	AAQAVN	SDQ	ESTT	ATAQ	PSASS	NTAA	353
		:	:	:	:	:	:	
Db	409	HVLA	-----	NOQLN	QIVH	QPSA	-----	459
		:	:	:	:	:	:	
Qy	354	GNGQWA	ASAAAA	GYRV	GSTP	SAGAVA	VWNDG	412
		:	:	:	:	:	:	
Db	460	GNGQWA	STAV	AKFG	KITV	PKVGN	AIACWS	519
		:	:	:	:	:	:	
Qy	413	SIGNR	GW	FNP	-----	GSVS	YIYP	432
		:	:	:	:	:	:	
Db	520	YISN	FRG	WFD	PTTSY	GLR	LTIY	544
		:	:	:	:	:	:	

```

RESULT 5
US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

```

```

QY      380  AVWNDCGYGHVAYVTGV-QCGQIQVQEAANYAGNQSIGNYRGWFPNGSVSYTY 430
      :      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      204  LOTSQGAIGHVAYVEGYNSNGSRVSENNYGHGAGVVTRTISASQAASYNY 255

RESULT 6
US-11-052-554A-3
; Sequence 3, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052.554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: PRRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-3

```

```

RESULT 7
US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227

```

```

; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226.5; DB 7; Length 394;
Best Local Similarity 27.5%; Pred. No. 1.le-06;
Matches 109; Conservative 61; Mismatches 178; Indels 49; Gaps 15;

QY   5 ILSAVLVSGVTLSSTLTSVAKADDFDAQIASQ-----DSKINNLTAQQQAQAQVNNTIQ 59
DB   16 ILSAVL--HVILPALIWS-----FDENIASAGGGSSIDAVWDSGAVVEQYKRMQ 68
QY   60 GOVSALQTQOAELOAENORLEAQSA TLGQQIQTLLSSKIVARNESLKQ-----QARSQKS 114
DB   69 SQESS-----AKRSDEQRKMKEQAAA-----EELFEKQAAEQERLKLKERLA AQOKK 118
QY   115 NNAATSVINAIINSKVS DAINRVSAIREVVSANEKMLHQODKAAVEQKHQENOAIAINT 174
DB   119 QAEEAAKQAE LKQKQAEAAKAAADAKAKAEADDKAAEEAAKKAADADAKKAAEA AAKA 178
QY   175 VVAANQETIAQNTNALNTQOALEAAQLNLQ---AELTTA QDQKATLVAQKAAAAEEAA--R 229
DB   179 ABAEQKKAEAAALUKKAAEAAEAABARKAAAEKAAADKKA---AEKAAAEKAAADK 235
QY   230 QAAAAQAAAEKAAAEKALOBQAQAQAQAQAANNNNNTQTATDASDOQAAAAADNNTQAATGDS 289
DB   236 KAAAEKAAADKKA-A-KAAAEKAAAKAA-----EADDIFGELS SGNAPKTGGG 286
QY   290 TQOSAAQAVNNSDQESTTATAAQ-PSASASTAAVAANTSSANTYPAGQCCTGWGVKSLAP- 347
DB   287 AKGNNASPAGSNTKNNGASGADINNYAGQIKSAIESKFYDASSYAGTKTCLTRIK-LAPD 345
QY   348 -WVGNYWGNGOWA-ASRAAGYRVGSTPSAGAVAVW 382
DB   346 GMLLDIKPEGDPALCQAAALAAKLA KIPKPPSQAVY 382

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip W
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; PRIOR FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

Query Match      10.6%; Score 224; DB 6; Length 1236;
Best Local Similarity 23.3%; Pred. No. 6e-06;
Matches 91; Conservative 89; Mismatches 185; Indels 26; Gaps 10;

```

Db 187 STYWSDAKYWASNAANDGYQVDNTPSVGAIMQSTPGPGYGHVAYVERINGDGSILISEMNY 246

QY 409 AGNQSIGNRY 418
| | | | |
Db 247 ANGPYNMYR 256

RESULT 10
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match 10.1%; Score 214; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSIQSAAQAVNNSDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345

Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDKT 67

QY 346 APWVGNYMGNGQWAAASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVQG-GQIQVQ 404

Db 68 GKGISTWGNANSWATAQAAGFTVNNTPBEGAIMQSSSEGAFGHVAFVESVNDGSIITS 127

QY 405 EANYAG 410
| | | | |
Db 128 ENNYDG 133

RESULT 11
US-10-793-626-2870
; Sequence 2870, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2870
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2870

Query Match 10.1%; Score 214; DB 6; Length 157;
Best Local Similarity 39.7%; Pred. No. 2.1e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSIQSAAQAVNNSDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345

Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDKT 67

QY 346 APWVGNYMGNGQWAAASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVQG-GQIQVQ 404

Db 68 GKGISTWGNANSWATAQAAGFTVNNTPBEGAIMQSSSEGAFGHVAFVESVNDGSIITS 127

QY 405 EANYAG 410
| | | | |
Db 128 ENNYDG 133

RESULT 12
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-23

Query Match 9.1%; Score 193.5; DB 6; Length 2101;
Best Local Similarity 24.1%; Pred. No. 0.00064;
Matches 76; Conservative 58; Mismatches 113; Indels 69; Gaps 10;

QY 2 KKRILSAVLVSGVTLSSTATLSA-----VKADDFDAQIASQDSKI-----NMLTAQQQAA 51

Db 405 KGEVLGDLVQLLETLKQEAATLAANNNTLOQARVEMLETERGQOEAKLLAERGHFEESKQOL 464

QY 52 QAVNTIQGVESALQTQQAEL-----QAENQRLAQSATLGGQIQIOTLSSKIVARNE---SL 104

Db 465 SSLITDLQSSINLSQAKEELEQASQAHGARLTAQVASLTSELTLTNATIQOQDQELAGL 524

QY 105 KQARSQAQKSNAAATSYINAINISKVSDAINRVSAIREVVSANEXMLHQEQDQKAAVEQK 164

Db 525 KQAAKEKQAQLAQT-----LQQEQASQGLRHQVQLSSSLKQK 563

QY 165 HOENQAAINTVAANQETTAQNTNALNTQQAQLAQAQLNLQAEELTTAQDQKATILVAQKAA 224

Db 564 EQQ-----LKEVAEKQEAATRD-----HAQQL-----ATAAEERASLRRERDAL 603

QY 225 EEAARQAAAAAQA 284

Db 604 KQL-----BALEKRAKLEILQOQLQVANEARDSAQTSTVQAREKABELSRKVEEL 655

QY 285 QTGDST---DQSAQA 297

Db 656 QACVETARQEQHEAQA 671

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.491 Seconds
(without alignments)
1668.609 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLGVGLTSSAT.....SIGNVRGMFPGSVIYIPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2105	100.0	431	7	ADD93653 Streptoco
2	2105	100.0	431	9	ADx37276 Streptoco
3	2096	99.6	431	7	ADD93650 Streptoco
4	2096	99.6	431	9	ADx37273 Streptoco
5	2092	99.4	431	7	ADD93649 Streptoco
6	2092	99.4	431	9	ADx37272 Streptoco
7	2092	99.4	431	9	ADx37272 Streptoco
8	2075.5	98.6	432	9	ADx37274 Streptoco
9	2075.5	98.6	432	9	ADx37274 Streptoco
10	2066.5	98.2	432	7	ADx37275 Streptoco
11	2066.5	98.2	432	9	ADx37275 Streptoco
12	1238	58.8	447	5	ABP29684 Streptoco
13	1238	58.8	447	8	ADU69524 S agalact
14	1238	58.8	447	8	ADV88392 Streptoco
15	1238	58.8	447	8	ADV81808 Streptoco
16	1238	58.8	447	8	ADv79645 Streptoco
17	1092.5	51.9	398	9	AEb91542 Microbial
18	1085.5	51.6	398	5	ABP25919 Streptoco
19	1085.5	51.6	398	8	ADP3884 S. pyogen
20	1080	51.3	395	5	ABP25918 Streptoco
21	1057.5	50.2	392	6	ABU02747 S. pneumo
22	1057.5	50.2	392	8	ADr50226 S_pneumon
23	1057.5	50.2	392	8	ADr50185 S_pneumon
24	1057.5	50.2	399	8	ADr94595 Novel S.

ALIGNMENTS

RESULT 1	
ADD93653	
ID	ADD93653 standard; protein; 431 AA.
XX	
AC	ADD93653;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Streptococcus mutans glucan binding protein-B.
XX	
KW	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX	
OS	Streptococcus mutans.
XX	
PN	WO2003075845-A2.
XX	
PD	18-SEP-2003.
XX	
PF	07-MAR-2003; 2003WO-US006962.
XX	
PR	07-MAR-2002; 2002US-0363209P.
PR	08-AUG-2002; 2002US-0402483P.
XX	
PA	(FORS-) FORSYTH INST.
XX	
PI	Smith DJ, Taubman MA;
XX	
DR	WPI; 2003-845091/78.
DR	GENBANK; AY046414.
XX	
PT	Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX	
PS	Claim 5; Page 8-9; 49pp; English.
XX	
CC	The present sequence is the protein sequence of the glucan binding protein-B (GbpB) of Streptococcus mutans strain 5S93. The sequence includes a number of human leucocyte antigen (HLA)-binding epitopes. The invention provides immunogenic compositions and vaccines for dental caries. The compositions comprise major histocompatibility complex (MHC) class II protein-binding GbpB peptides, especially HLA-binding peptides, covalently linked with peptide subunits of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in

```

CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match      100.0%; Score 2105; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 7e-134;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAEIQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
Db 61 QVSALQTOQAEIQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAREVVSANEMKLOQEQDKAAVEQKQENQAINTVAAQOE 180
Db 121 INAIINSKVSDAINRVSAREVVSANEMKLOQEQDKAAVEQKQENQAINTVAAQOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDKATLVAQKAAAEAEARQAAAAQAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDKATLVAQKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
QY 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 2
ADX37276
ID ADX37276 standard; protein; 431 AA.
XX
AC ADX37276;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #5.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
FN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;

WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 33; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 431 AA;

Query Match      100.0%; Score 2105; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 7e-134;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAEIQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
Db 61 QVSALQTOQAEIQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAREVVSANEMKLOQEQDKAAVEQKQENQAINTVAAQOE 180
Db 121 INAIINSKVSDAINRVSAREVVSANEMKLOQEQDKAAVEQKQENQAINTVAAQOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDKATLVAQKAAAEAEARQAAAAQAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDKATLVAQKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
QY 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 3
ADD93650
ID ADD93650 standard; protein; 431 AA.
XX
AC ADD93650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
FN WO2003075845-A2.
XX
PD 18-SEP-2003.

```

ADX37273 standard; protein; 431 AA.

ADX37273;

21-APR-2005 (first entry)

Streptococcus mutant glucan binding protein B variant #2.

immunogenicity; immune stimulation; glucan binding protein-B; microparticle; major histocompatibility complex; tooth disease.

Streptococcus mutans.

US2005031633-A1.

10-FEB-2005.

09-MAR-2004; 2004US-00797821.

13-APR-1998; 98US-0081550P.

08-JAN-1999; 99US-0115142P.

12-APR-1999; 99US-00290049.

07-MAR-2002; 2002US-0363209P.

08-AUG-2002; 2002US-0402483P.

07-MAR-2003; 2003US-00383930.

(SMIT//) SMITH D J.

(TAUB//) TAUBMAN M A.

Smith DJ, Taubman MA;

WPI; 2005-151644/16.

New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.

Claim 3; SEQ ID NO 30; 73pp; English.

The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.

Sequence 431 AA;

Query Match 99.6%; Score 2096; DB 9; Length 431;

Best Local Similarity 99.5%; Pred. No. 2.8e-133;

Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQG 60

DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQG 60

QY 61 QVSALQTQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQARSQAQKNAATSY 120

DB 61 QVSALQTQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQARSQAQKNAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDQKAAVEQKQENQAQVNTVAANQE 180

DB 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDQKAAVEQKQENQAQVNTVAANQE 180

QY 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAAARQAAAAQAAAA 240

DB 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAAARQAAAAQAAAA 240

QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAQAQVNN 300

DB 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAQAQVNN 300

RESULT 4

ADX37273

```
QY 301 DQESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGQWAA 360
Db |||||||
QY 301 DQESTTATEAQPSSASSASTAAVANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGQWAA 360
Db |||||||
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
Db |||||||
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
Db |||||||
QY 421 NPGSVSYIYPN 431
Db |||||||
QY 421 NPGSVSYIYPN 431
Db |||||||
```

RESULT 5

```
ADD93649
ID ADD93649 standard; protein; 431 AA.
```

XX AC

ADD93649;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucan binding protein-B.

XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus mutans.

FH Key Location/Qualifiers

```
FT Region 6..25 /note= "HLA-binding peptide"
FT Region 16..35 /note= "HLA-binding peptide"
FT Region 33..52 /note= "HLA-binding peptide"
FT Region 37..56 /note= "HLA-binding peptide"
FT Region 48..67 /note= "HLA-binding peptide"
FT Region 52..71 /note= "HLA-binding peptide"
FT Region 88..107 /note= "HLA-binding peptide"
FT Region 113..132 /note= "HLA-binding peptide"
FT Region 117..136 /note= "HLA-binding peptide"
FT Region 137..156 /note= "HLA-binding peptide"
FT Region 174..193 /note= "HLA-binding peptide"
FT Region 194..213 /note= "HLA-binding peptide"
FT Region 214..233 /note= "HLA-binding peptide"
FT Region 248..267 /note= "HLA-binding peptide"
FT Region 289..308 /note= "HLA-binding peptide"
FT Region 306..325 /note= "HLA-binding peptide"
FT Region 311..330 /note= "HLA-binding peptide"
FT Region 349..368 /note= "HLA-binding peptide"
FT Region 365..384 /note= "HLA-binding peptide"
FT Region 383..402 /note= "HLA-binding peptide"
FT Region 403..422 /note= "HLA-binding peptide"
```

XX PN WO2003075845-A2.

XX RESULT 6

```
PD 18-SEP-2003.
XX PF 07-MAR-2003; 2003WO-US0006962.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PA (FORS-) FORSYTH INST.
XX PI Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
XX GENBANK; AY046410.
XX PT Composition useful as vaccines for dental caries comprises a fragment of
XX PT a glucan binding protein-B binding to a major histocompatibility complex
XX PS class II protein.
XX PS Claim 6; Page 7; 49pp; English.
XX CC The present sequence is the protein sequence of the glucan binding
XX CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence
XX CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
XX CC invention provides immunogenic compositions and vaccines for dental
XX CC caries. The compositions comprise major histocompatibility complex (MHC)
XX CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
XX CC covalently linked with peptide subunits of a glucosyltransferase. The
XX CC compositions are used in a claimed method of eliciting production of an
XX CC antibody in a mammal. Dieptopic or multiepitopic polypeptides can be
XX CC prepared synthetically or by recombinant DNA technology. Antibodies
XX CC raised against MHC class II binding fragments of GbpB can be used in
XX CC passive immunisation.
XX SQ Sequence 431 AA;
Query Match 99.4%; Score 2092; DB 7; Length 431;
Best Local Similarity 99.3%; Pred. No. 5.3e-133;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQG 60
Db |||||||
QY 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQG 60
Db |||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAKSNAATSY 120
Db |||||||
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAKSNAATSY 120
Db |||||||
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQSQDQKAAVEQKQENQAINTVAANQE 180
Db |||||||
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQSQDQKAAVEQKQENQAINTVAANQE 180
Db |||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEFLTADOKATLVAQKAAAEAAQAAQAAEA 240
Db |||||||
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEFLTADOKATLVAQKAAAEAAQAAQAAEA 240
Db |||||||
QY 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQQAADANTQAAQTGDSQSAQAQVNNNS 300
Db |||||||
QY 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQQAADANTQAAQTGDSQSAQAQVNNNS 300
Db |||||||
QY 301 DQESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGQWAA 360
Db |||||||
QY 301 DQESTTATEAQPSSASSASTAAVANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGQWAA 360
Db |||||||
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
Db |||||||
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
Db |||||||
QY 421 NPGSVSYIYPN 431
Db |||||||
QY 421 NPGSVSYIYPN 431
Db |||||||
```


CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from
 CC distantly related organisms, and from bacteria belonging to a wide
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
 CC unique proteins. The present sequence is a microbial pathogen adhesin
 CC protein sequence.
 XX
 SQ Sequence 431 AA;

Query Match 99.4%; Score 2092; DB 9; Length 431;
 Best Local Similarity 99.3%; Pred. No. 5.3e-133;
 Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTTIG 60
 DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTTIG 60
 QY 61 QVSALQTOQAELOAENQRLAQSAATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 DB 61 QVSALQTOQAELOAENQRLAQSAATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 QY 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180
 DB 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180
 QY 181 TTAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAAAE 240
 DB 181 TTAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAAAE 240
 QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDSDQQAADNTQAQTDGSTEQSAQAQAVNNS 300
 DB 241 KAAAEKALQEQAAQAAQAAANNNTQATDSDQQAADNTQAQTDGSTEQSAQAQAVNNS 300
 QY 301 DOESTTATEAQPSSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360
 DB 301 DOESTTATEAQPSSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360
 QY 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
 DB 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
 QY 421 NPGSVSYIYPN 431
 DB 421 NPGSVSYIYPN 431

RESULT 8
 ADD93651
 ID ADD93651 standard; protein; 432 AA.
 XX AC ADD93651;
 XX 29-JAN-2004 (first entry)
 DE Streptococcus mutans glucan binding protein-B.
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
 OS Streptococcus mutans.
 XX WO2003075845-A2.
 XX 18-SEP-2003.
 XX 07-MAR-2003; 2003WO-US006962.
 XX 07-MAR-2002; 2002US-0363209P.
 XX 08-AUG-2002; 2002US-0402483P.
 XX (FORS-) FORSYTH INST.
 XX Smith DJ, Taubman MA;
 XX WPI; 2003-845091/78.

DR GENBANK; AY046412.
 XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.
 XX
 PS Claim 5; Page 8; 49pp; English.
 XX
 CC The present sequence is the protein sequence of the glucan binding
 CC protein-B (GbpB) of Streptococcus mutans strain 150P2. The sequence
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
 CC invention provides immunogenic compositions and vaccines for dental
 CC caries. The compositions comprise major histocompatibility complex (MHC)
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
 CC covalently linked with peptide subunits of a glucosyltransferase. The
 CC compositions are used in a claimed method of eliciting production of an
 CC antibody in a mammal. Pieptopic or multiptopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.
 XX
 SQ Sequence 432 AA;

Query Match 98.6%; Score 2075.5; DB 7; Length 432;
 Best Local Similarity 98.6%; Pred. No. 6.9e-132;
 Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
 QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTTIG 60
 DB 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAQOQAAQAVNTTIG 60
 QY 61 QVSALQTOQAELOAENQRLAQSAATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 DB 61 QVSALQTOQAELOAENQRLAQSAATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 QY 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180
 DB 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180
 QY 181 TTAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAAAE 240
 DB 181 TTAQNTNALNTQQAQLEAAQNLQALHTTAQDKATLVAQKAAAEAEARQAAAAQAAAE 240
 QY 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDSDQQAADNTQAQTDGSTEQSAQAQAVNN 299
 DB 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDSDQQAADNTQAQTDGSTEQSAQAQAVNN 300
 QY 300 SDQESTTATEAQPSSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWA 359
 DB 301 SDQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWA 360
 QY 360 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGQGGQIQVQEAANYAGNQSIGNYRGW 419
 DB 361 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGQGGQIQVQEAANYAGNQSIGNYRGW 420
 QY 420 NPGSVSYIYPN 431
 DB 421 NPGSVSYIYPN 432

RESULT 9
 ADX37274
 ID ADX37274 standard; protein; 432 AA.
 XX AC ADX37274;
 XX 21-APR-2005 (first entry)
 XX Streptococcus mutant glucan binding protein B variant #3.
 DE immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX

Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQOEADKALEAKQIENQNAINTVAANKQ 177
Qy 181 TTAQNTNALNTQQAQLEAAQLNLQAELETTAQQOKATLVAQKAAABEAARQAAAAQAABA 240
Db 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAABA 237
Qy 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTE--QSAAQAVN 298
Db 238 KAQAQAKAQABSAVAKAQA---QVESATAPTETVQTPRTEIKPSNLTATS 286
Qy 299 NSDQESTTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336
Db 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVRSYDSSNTYPMQ 346
Qy 337 CTWGVKSLAPWGVNNGWGQWAAASAAAGYRVGSTPSAGAVVM--NDGGYGHVAYVTG 394
Db 347 CTWGAKSMAWGVNNGWGNQWGSARAAAGYSGVTTPRVGAVMWPDYDGGYGHVAVVTS 406
Qy 395 V-QGGQIQVQEAANYAGNOSIGNYRGWFPN--GSVSYIYPN 431
Db 407 VANNSSIQVMESNYAGNMSIGNYRGSFNPSAGSGSVYIYPN 447

RESULT 14
ADV88392
ID ADV88392 standard; protein; 447 AA.
AC ADV88392;
XX
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae protein sequence, SEQ ID 786.
XX
XX Antibacterial; Vaccine; bacterial infection.
XX
XX Streptococcus agalactiae.
XX
XX FR2824074-A1.
XX
XX 31-OCT-2002.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP) INST PASTEUR.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 6; SEQ ID NO 786; 2687pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
XX ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
XX agalactiae involved in the synthesis of amino acids, cell membranes, of
XX intermediate (central) metabolism, energetic metabolism including purines,
XX pyrimidines and/or nucleosides, nucleotide metabolism including purines,
XX transcription, translation, protein transport, adaptation to atypical
XX conditions, sensitivity to medicines and/or analogues, functions related
XX to transposons, biosynthesis of cofactors, prosthetic groups and
XX transporters, cell membrane proteins and cellular machinery. (I) are
XX useful for the detection and/or amplification of nucleic acids.
XX Pharmacological composition comprising (I) or (II) are useful for
XX treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2
CC contains 6617 sequence whereas the present patent only contains 2344
CC sequences.
XX
SQ Sequence 447 AA;
Query Match 58.8%; Score 1238; DB 8; Length 447;
Best Local Similarity 60.7%; Pred No. 2, 7e-75;
Matches 280; Conservative 43; Mismatches 94; Indels 44; Gaps 10;
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQTASQDSKINNLTAAQOAAQAVNTIAG 60
Db 1 MKKRILSAVLVSGVTLTGTAAV--TVNADDFDKIAATDSVINTLSGQOAAAQNVTAIKG 58
Qy 61 QVSALOTOQAELOAENQRLAQSATLQGOIOTLSSKIIVARNESLKQOARSQAQSNATSY 120
Db 59 QVGALESQOQSELAQNAQLEAVSQQLGQEIOTLSNKKIIVARNESLKKQVRSQAQGN-LTNY 117
Qy 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQOQOQDQKAAVEQKQENQAAINTVAANO 180
Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQOEADKALEAKQIENQNAINTVAANKQ 177
Qy 181 TTAQNTNALNTQQAQLEAAQLNLQAELETTAQQOKATLVAQKAAABEAARQAAAAQAABA 240
Db 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAABA 237
Qy 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTE--QSAAQAVN 298
Db 238 KAQAQAKAQABSAVAKAQA---QVESATAPTETVQTPRTEIKPSNLTATS 286
Qy 299 NSDQESTTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336
Db 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVRSYDSSNTYPMQ 346
Qy 337 CTWGVKSLAPWGVNNGWGQWAAASAAAGYRVGSTPSAGAVVM--NDGGYGHVAYVTG 394
Db 347 CTWGAKSMAWGVNNGWGNQWGSARAAAGYSGVTTPRVGAVMWPDYDGGYGHVAVVTS 406
Qy 395 V-QGGQIQVQEAANYAGNOSIGNYRGWFPN--GSVSYIYPN 431
Db 407 VANNSSIQVMESNYAGNMSIGNYRGSFNPSAGSGSVYIYPN 447

RESULT 15
ADV81808
ID ADV81808 standard; protein; 447 AA.
XX
XX ADV81808;
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae protein, SEQ ID 2949.
XX
XX Antibacterial; vaccine; bacterial infection.
XX
XX Streptococcus agalactiae.
XX
XX WO200292818-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-IB003059.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX

Search completed: February 15, 2006, 18:03:01
Job time : 114.491 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:03:21 ; Search time 20.9676 Seconds
(without alignments)
1977.789 Million cell updates/sec

Title: US-10-797-821-33
Perfect score: 2105
Sequence: 1 MKKRILSAVLVSGVTLSSTAT.....SIGNVRGMFPGSVSYIYPN 431
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1057.5	50.2	392	2 G95258	secreted 45 kd pro
2	1057.5	50.2	392	2 B98124	general stress pro
3	672	32.0	461	2 JN0097	secreted 45K prote
4	661.5	31.4	456	2 E89903	hypothetical prote
5	551	26.2	507	2 S05542	hypothetical prote
6	359.5	17.1	398	2 AC1763	peptidoglycan lyti
7	346	16.4	401	2 A11387	peptidoglycan lyti
8	319.5	15.2	473	2 F70031	cell wall-binding
9	266	12.6	581	2 E75383	conserved hypotet
10	259	12.3	436	2 AHI387	cell wall binding
11	257.5	12.2	461	2 H84099	cell wall-binding
12	256.5	12.2	437	2 AB1763	cell wall binding
13	238.5	11.3	528	2 B75310	conserved hypotet
14	230.5	11.0	421	2 JV0057	tola protein - Esc
15	227	10.8	971	2 B90835	probable tail fibe
16	227	10.8	973	2 C85693	probable membrane
17	226	10.7	394	2 F90725	membrane spanning
18	226	10.7	394	2 G85576	membrane spanning
19	225.5	10.7	166	2 C90029	hypothetical prote
20	223.5	10.6	267	2 F90028	hypothetical prote
21	222.5	10.6	265	2 B98937	hypothetical prote
22	221.5	10.5	392	2 F96937	cell wall-associat
23	221.5	10.5	492	2 A28616	cell wall surface
24	220.5	10.5	4776	2 E95206	cell wall surface
25	220	10.5	255	2 G90061	hypothetical prote
26	217.5	10.3	1122	2 G64887	probable tail fibe
27	215	10.2	688	2 A83179	conserved hypotet
28	214.5	10.2	536	2 D84325	Htr17 transducer l
29	212.5	10.1	483	2 A26297	M6 protein - Strep

RESULT 1
G95258
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: G95258
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95258
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-392 <KUR>
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE005672; I
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2216

Query Match 50.2%; Score 1057.5; DB 2; Length 392;
Best Local Similarity 50.6%; Pred. No. 2e-44;
Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

Qy	1	MKKRILSAVLVSGVTLSSTATILSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60
Db	1	MKKRILASLLSTVMVSVQAVLTTHAETDDKIAAQNKNISLNTAAQQAQAQVNDIQE	60
Qy	61	QVSALQTOQAEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQSNAAATSY	120
Db	61	QVSATQAEQSNLQAEQNDRLQAESKKEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY	120
Qy	121	INAIINSSVSDAIRNVSAREVWSANERKMLQQQEQDQKAAVFEQKQEQENQAQINTVAANQE	180
Db	121	INTIVNSKITEAISRKVAAMSIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQQ	180
Qy	181	TTAQNVTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAKAAAEAAQAAAAQAAEA	240
Db	181	KLADDAQATTTKQAEKAAELSLAAEKATAEGEKASLLEQKAAAEAAEAFAVAEAAAYK	239
Qy	241	KAAAEKALQEAQAQAQAAAAANNNTQATDASDQAAAAADNTQAATQDGTSTEQSAQAQVNN	300
Db	240	-----EKRAEQQQSVLASANTNLTAQ-----VQAVSES	267
Qy	301	DOESTTATEAOPSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWPVGNWNGGQWAA	360
Db	268	-----AAAPVRKVRPTYSTNASSYPIGECTGWGKTLAPWAGDNGNGAOWAT	315
Qy	361	SAAAAGYRVGSTPSPAGAVAVNDGGYGHVAYVTGVQG-GQIQVEANYAGNOSIGNYRGW	419

ALIGNMENTS

Db 316 SAAAAGFRGTSPQVGAIACMNDGGYGHVAVVTAVESTTTRIQVSESNYAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVTIY 390

RESULT 2

B98124

General stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004

C:Accession: B98124

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; Eyer, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.B.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; et al.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; PMID:21429245; PMID:11544234

A:Accession: B98124

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317; C:Genetics:

A:Gene: gsp-781

Query Match 50.2%; Score 1057.5; DB 2; Length 392;

Best Local Similarity 50.6%; Pred. No. 2e-44;

Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60

Db 1 MKKKILASLLSTVWVSQVAVLTTHAETTTDDKIAAQDNKI SNLTAAQOQEAQKQVDIOE 60

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

Db 61 QVSAIQAEOSNQAENDRIQAESKKLEGEITELSKNIVSRNOSLEKQAQSAQTNGAVTSY 120

QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOQODKAAVPOKQOENQAAINTVAANOE 180

Db 121 INTIVNSKITEAISRVAAWSIVSANNKMLQEQADKKAISEKVANNDAINTVIANQO 180

QY 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQOKATLVAQKAAAEARQAQAAAAQAAEA 240

Db 181 KLADDAQALTTKQAEKKAELSAAEKATPAEGEKASLLEQKAAAEARAAAFAVAEAAVK- 239

QY 241 KAAAEKALQEAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAQAQVNN 300

Db 240 -----EKRAQQQSIVLASANTNLTAAQ-----VQAVSES 267

QY 301 DOESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTGWKSLAPWVGNVWNGGQWAA 360

Db 268 -----AAAPVRKVRPTVSTNASSYPIGECTVGTKLAPWAGDYWGNGAQWAT 315

QY 361 SAAAAGRVGSPSPSAGAVVNDGGYGHVAVVTGVOG-QGIQVQEAANYAGNOSIGNYRGW 419

Db 316 SAAAAGFRGTSPQVGAIACMNDGGYGHVAVVTAVESTTTRIQVSESNYAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVTIY 390

RESULT 3

JN0097

secreted 45K protein precursor - Lactococcus lactis

C:Species: Lactococcus lactis

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: JN0097

R:van Asseldonk, M.; Rutten, G.; Otman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.

Gene 95, 155-160, 1990

A:Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis sut

A:Reference number: JN0097; PMID:91071599; PMID:2123812

A:Accession: JN0097

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <VAN>

A:Cross-references: UNIPARC:UPI000016D72B; GB:M35374

Query Match 32.0%; Score 673; DB 2; Length 461;

Best Local Similarity 35.2%; Pred. No. 8.6e-26;

Matches 169; Conservative 104; Mismatches 139; Indels 68; Gaps 13;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60

Db 1 MKKKILASLLSTVLSRAAPLSGVYAT-NSDIAQADATISSAKAQAQAQVDSLOS 59

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

Db 60 KVDLSLQKQKTSKQAIKIESEAKALNAQIATLINESIKERTKTLEAQAARSQVNSSATNY 119

QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOQODKAAVPOKQOENQAAINTVAANOE 180

Db 120 MDAVVNSKSLTDVIQKVTATIAITVSSANKMLQEQEKELQKSETVKCKYNNQVFLSQ 179

QY 181 TTAQNTNALNTQOALEAAQLNLQAELTAAQOKATLVAQKAAAEARQAQAAAAQAAEA 240

Db 180 SLDQAQELTQQAEKLVATLNYQATIAQKQALLDEKAAAEKAAQEAQKQAYEA 239

QY 241 KAAAEKALQEAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQ-----SAAQ 295

Db 240 Q-----QKEAAQAQAASTAATAKAVEAATSSASASSQAPOVSTSTDTNTTSNAS 291

QY 296 AVNNSDQESTTATEAQPSSASSASTAVVTANTSS-----ANTYP 333

Db 292 SSNSSSSSSSSSSSSSSSSSSSSSNAGGNTNSTGTGNTGTTGGSGINSSPIGNPYA 351

QY 334 AGOCT---WG-----VKSLAPWVGNVWNGGQWAAASAAAAG--YRVGSPSPSAGAVAV 380

Db 352 GGGCTDYWQVFAAGCIYIRIMP-----GNGGQWASNGPAQGVLVHVGAAP--GVTA 403

QY 381 -----WPDGGYGHVAVVTGVOG-QGIQVQEAANYAGNOSIGNYRMFNPQSVSYIYPN 431

Db 404 SPSADFDVGANSPYGHVAIVKSVNSDGTITIKEGY-GTTWGHGR-TVSGASGVTLFMPN 461

RESULT 4

E86903

hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004

C:Accession: E86903

R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi

A:Reference number: A86625; PMID:21235186; PMID:11337471

A:Accession: E86903

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-456 <STO>

A:Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI00000C6BFC; GB:AE005176; PID:g12725296; P1

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: usp45

Query Match 31.4%; Score 661.5; DB 2; Length 456;

Best Local Similarity 35.4%; Pred. No. 3e-25;

Matches 169; Conservative 103; Mismatches 138; Indels 67; Gaps 15;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60

Db 1 MKKKILASLLSTVLSRAAPLSGVYAT-NSDIAQADATISSAKAQAQAQVDSLOS 59

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

Db 60 KVDLSQKQASTKAQIAKIESEAKALNAQIATLNESIAERTTKTLEAQASQVNSSATY 119
Qy 121 INAINSKSVSDAINRVSAIREVVSANERKMLQOEODKAAVEQKQOENQAAINTVAANOE 180
Db 120 MDAVNSKSLTDVIOKVTAIATVSSANKQMLQOEKEQELQKQSEKQVTKYNQNFVLSQ 179
Qy 181 TTAQNTNALNTQQAQLEAAQLNLQALBELTTAQDQKATLVAKKAAAEAEARQAAAAQAAEA 240
Db 180 SLDSSQAQELTSQAELKVATLNTQATIAATQDKQSLDLDEKAAAEKAAQEAQKQAAVEA 239
Qy 241 KAAAEKALQEAQAQAQAANNTQATDASDQQAARAADNTQAAQTGDSQTEQ-SAAQAVNN 299
Db 240 Q-----QKEAAKAQAATAAT--TKAVEEATSTVSSQASQSSSSSTNTSSNTSN 289
Qy 300 SQUESTTATEAOPSSAST-----AVVTANTSS-----ANTYPACQ 336
Db 290 SSSSSSSSSSSSSSSSSSGGNTGNNNAAGTGTGSSSSGINSITPIANPYAGGG 349
Qy 337 CT---WG-----VKSLAPWVGNWNGGQWAAASAAAG--YRVGSTPSSAGAVAV--- 380
Db 350 CTDYVQWYPAAQGIYIRNMP-----GNGGQWATNGPAQGVHLHVGAAP--GVIASPSS 401
Qy 381 -----WNDGGYGHVAVYTCV-QGGQIQVQEAANYAGNQSIGNVRGWFNPGSVSIYVN 431
Db 402 ADFVGYANSYPYGHVAIVKSVNSNGTITIKEGGY-GTTWGHGR-TVSAGVTFLLMPN 456

RESULT 5
S05542
Hypothetical protein, 54K - Enterococcus faecium
C:Species: Enterococcus faecium
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S05542
R:Fuerst, P.; Moesch, H.U.; Solioz, M.
Nucleic Acids Res. 17, 6724, 1989
A:Title: A protein of unusual composition from Enterococcus faecium.
A:Reference number: S05542; MUID:89385998; PMID:2780297
A:Accession: S05542
A:Molecule type: DNA
A:Residues: 1-507 <FUE>
A:Cross-references: UNIPROT:P13692; UNIPARC:UPI000016F6RC; GB:X16421; EMBL:M26048; NID:5
A>Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 26.2%; Score 551; DB 2; Length 507;
Best Local Similarity 28.9%; Pred. No. 7.3e-20;
Matches 147; Conservative 93; Mismatches 166; Indels 102; Gaps 13;

Qy 10 LVSGVTLSSATLSAVKADDFQAQIASQDSKINNLTAAQQAQAQVNTIQQVSA 69
Db 1 MLSSIALTAVGSPFAAAADDFDSIQIQQDKIADLQNAQSAQSQIEALEGQVSAINTKA 60
Qy 70 ABLQENQBLEAQSATLGQIQITLSKIVARNESLKQARSQAQKSNAAATSYNAIINSKS 129
Db 61 QDLTKQDTLRKESAQKQEIKDQRIEKEATIKQARETVQVNTSSNYIDAVLNADS 120
Qy 130 VSDAINRVSAIREVVSANERKMLQOEODKAAVEQKQOENQAAINTVAANQETIAQNTAL 189
Db 121 LADAVGRIQAMTIVKANQDLVQQEKEDQAVEAKAEAKQELADNQAALESQKGLD 180
Qy 190 NTQQAQLEAAQLNLQALBELTTAQDQKATLVAKKAAAEAEARQAAAAQAAQ 235
Db 181 LAKQADNLVLTSLAAEQATAEDKKADLNKKAEEAEQARIEQARLAEQARQAAQAEK 240
Qy 236 AAEEKAAAEAKALQEQQAQAQAANNNTQATDASDQQAARAADN-----TQ 281
Db 241 AEKARQAAAAQAQQAQALSSASTTTTSSSAQSSSEESKAPESSTTESTSTSTTTE 300
Qy 282 AAQTGDSQTEQAAA---QAVNNSDQESTTATEAQPSSASTAV-----VT 323
Db 301 NSTGSSSTESSTESTVPESTQESTPANTESSSSSTNTVNNNTNNSTNNSTNT 360
Qy 324 ANTSSANTY-----PAGQCTWGVKSLA-----PMVGNWNGNG----- 355

Db 361 NNNNNNTVTPAPTPTPTPPAPAPNPSPGVSNGAAIIVAEAYKYIGTPYV---WGGKDPSPG 417
Qy 356 -----GOWAASAAAAGYRVG-STPSAGAVAVVWMD--GGYGHVAVYVTG 394
Db 418 PDCSGFTRVYLVQVTRDIGGTVTPQESAGTKLSVSAQAGDLLFWGSAGTYHVAISLG 477
Qy 395 VGGQIQVQEAANYAGNQSIGNVRGWFNP 422
Db 478 --GGQ-YIHAPQPGENVKVGVSQ-WYTP 501

RESULT 6
AC1763
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clp11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1763
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <GLA>
A:Cross-references: UNIPROT:Q92Y78; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:CAC97875.1;
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: spl

Query Match 17.1%; Score 359.5; DB 2; Length 398;
Best Local Similarity 27.0%; Pred. No. 9.7e-11;
Matches 113; Conservative 89; Mismatches 146; Indels 71; Gaps 16;

Qy 1 MKKRILSAV-LVSGVTLSSATLSAVKADDFQAQIASQDSKINNLTAAQQAQAQVNTIQ 59
Db 1 MKNTFIALSLAAVLSLTPAFTNVPA--DVNTDQNDQDKINDIKSKTKGLQSDLSLV 58
Qy 60 GOVSALQTOQAELQAEQNRLEAQSATLGQIQITLSKIVARNESLKQARSQAQKSNAAATSY 119
Db 59 ADLEKAQEKAKSLQGEFDQTKELQNLQDIKINERIKERETVLKERAMQKTSNSNA 118
Qy 120 YINALINSKVSDAINRVSAIREVVSANERKMLQOEOD-----KAAVEQKQOENQAAI 172
Db 119 YLEVLDAENLSDLVGRVSAVNQLVDSKSILEDQDKKALKTKQTAVKKQEEQATAI 178
Qy 173 NTVAAANQETIAQNTNALNTQQAQLEAAQLNLQALBELTTAQDQKATLVAKKAAAEAEARQAA 232
Db 179 HEFEAAQ-----NKIEAQAEKAEIIVAQLAADQASAEKAGLVSE---DKAAKE-- 226
Qy 233 AAQAAAEAKAAAEKALQEQQAQAQAANNNTQATDASDQQAARAADNTQAAQTGDSQTEOS 292
Db 227 -----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTSN 259
Qy 293 AAQAVNNSDQESTTATEAOPSSASTAVVTANTSSANTYPAG-----QCTWGVKSL- 344
Db 260 KYESTNNSEAPSA----ATPSSGGYSAMITAAARAQLGKPYSLGATGPSAFDCS-GFTSYA 314
Qy 345 --APWGNVGNWNGGQWAAASAAAAGYRVGSTPSAGAVAVVWMD--GGYGHVAVYVTGQGGQI 400
Db 315 FRAAGVSLPRISGGYAAASKIS----ASQAEFGDLVFFNYGGIAHVGIYVG--GGQM 367

RESULT 7
A11387
peptidoglycan lytic protein P45 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: A11387
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecher, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.; Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <GLA>
A:Cross-references: UNIPROT:Q9R04; UNIPARC:UPI00000D019D; GB:NC_003210; PIDN:CAD00583.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: spl

Query Match 16.4%; Score 346; DB 2; Length 401;
Best Local Similarity 26.3%; Pred. No. 4.4e-10;
Matches 110; Conservative 93; Mismatches 148; Indels 68; Gaps 15;

QY 1 MKKRILSAVLVSGVTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 59
DB 1 MKKNTFIAISLAAVISLTPAFTNVFA--DVNTDIQNQDKINDIKSKKTDLSGLV 58

QY 60 GOVSALOTQQAELQENQLEAQSITLGGQIQTLSKIVARNESLKOQARSQAQNSAATS 119
DB 59 ADLEKAEQAKSLOGEPFKTGKELKLNEDIKSINERIKERTVLKERARAMQKTSNSNA 118

QY 120 YINAIIINSKVSDAINRVSAIREVVSANEKMLQOQEQDKAAVEQKQ-----QENQA-AI 172
DB 119 YLEVLIDNENLSDLVGRVSAVNVQLVDSKSLLEDQONDEKALKTKQTAVKKQEQEQATAI 178

QY 173 NTVAAQNQETIAQNTNALNTQQAELAAQLNLAELTTAQQDKATLVAQKAAAEAAQAA 232
DB 179 HEYEAQQ-----NKIEAQKAEKAIVAQLASDQASAEAKAGLVSE--DKAAKE-- 226

QY 233 AAQAAAEKAAAEKALQEAQAAQAAANNNTQATDASDQQAADNTQAAQTGSTQOS 292
DB 227 -----ATARAVALRE-----ATSSNVQBSSTSTPSKSKNTTKN 263

QY 293 AAQAVNNSDOESTTATEAOPSASSASTAVVTANTSSANTYPAG-----CCTGWGKSL- 344
DB 264 VASNDNNSAPSAT-----PSSGGYSAMISAANAQLGKPYSLGATGPSAFDCS-GFTSYA 317

QY 345 --APWGVNMGNGQMAASAAAAGYRVGSTPSAGAVAVVNDG-GYGHVAVYTVGQGGQI 400
DB 318 FRAAGVSLPRTSGGQYAAASKIS---ASQAKPGDLVFFNYGSGIAHVGIYVG--GGQM 370

RESULT 8
F70031
cell wall-binding protein homolog yvce - *Bacillus subtilis*
C:Species: *Bacillus subtilis*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: F70031
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler, ich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, C.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumetate, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F70031
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-473 <KUN>
A:Cross-references: UNIPROT:P40767; UNIPARC:UPI0000060AEL; GB:Z99121; GB:AL009126; NID:g
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvce

Query Match 15.2%; Score 319.5; DB 2; Length 473;
Best Local Similarity 24.0%; Pred. No. 9.9e-09;
Matches 111; Conservative 95; Mismatches 186; Indels 71; Gaps 11;

QY 1 MKKRILSAVLVSGVTLS-----ATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQ 54
DB 1 MKKSLITGLASVIGTSTFLPFTSKTASAEITLDEKKKIESKQSEV---ASSIEAKEKE 57

QY 55 VNTIQGVSAQTQQAEL-----QAEENQLEAQSATLGGQIQTLSKIVAR 100
DB 58 LTELQENQSKIEKELKDINDKALDTSNKIEDKKEENDTKBEIKKLKKEIKETEAREIEKR 117

QY 101 NESLKOQARSQAQNSAATSINAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDKAA 160
DB 118 NEILKRVSLQESGSGQYIDVLGSGTSFGDFISRTAVSSIVDADKDLIKQEQDKAK 177

QY 161 VE-----QKQENQAAINTVAANQETIAQNTNALNTQQAELAAQLNLAELTTAQQD 213
DB 178 LEDSEADLNKLEKQVAAALAKLETWKQDLQOLNEKQKLFDEAKASQKTKAKAISLSE 237

QY 214 KATLVAQKAAAEKALQEAQAAAEKALQEAQAAQAAANNNTQATDASDQ 273
DB 238 ASELANQKANT-----AEQARIKKEQEAALIKKQEEAOKA-----SDETQTDDSQ 285

QY 274 AAAADNTQAAQTGSTQEOSAAQAVNNSQESTTATEAOPSASSASTAVVTANTSSANTYP 333
DB 286 TATTESSKASSDDSDNSDSSNGSSNSNGSSSKSGKSGNSNGSGTVTSNNGGIEG 345

QY 334 AGQCTWGVKSLAPVGVNMGNG-----OWA-ASAAAAGYRVGSTPSA 375
DB 346 AIVSGSSIVGGSP-----YKFGGRTQSDINNRI FDCSFVRWAYASAGVNLGPGVGTITD 401

QY 376 GAVAVVNDGGYHVAVYTVGQGGQIQVQEAANYAGNQSIGNYRG 418
DB 402 TLV-----GRQAVSASEMKRGDL-VFFDITYKTNHGVGIYLG 437

RESULT 9
E75383
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: E75383
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75383
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-581 <WHI>
A:Cross-references: UNIPROT:Q9RU45; UNIPARC:UPI00000C195A; GB:AE001998; GB:AE000513; NID:
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1549
A:Map position: 1

Query Match 12.6%; Score 266; DB 2; Length 581;
Best Local Similarity 23.5%; Pred. No. 4.7e-06;
Matches 122; Conservative 75; Mismatches 173; Indels 150; Gaps 17;

QY 16 LSSATTL SAVKADDFDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQG 61


```
Db 40 LAGGDTLSTRLEQLQRIEQRRLSVQKQKRLQTEVRARLRLQLNAQRAALDRDLALTSE 99
Qy 62 VSALOTQQAELQAEORLEAQAATLQ---QIQTLSKIVARNESLKQQAASAKS---N 115
Db 100 VTDLENELADVLA---RVTAATPALRETEAQRIVTSQV-----EALKVDARAVMKALYRA 152
Qy 116 AATSYNAIINSGSVSDAINRVSAIREVVSANEKMLQ-----QEODKAAY 161
Db 153 RNTQVNRLLSQNSISDMLIRLDYANNAGORNVVMEQLRGAAGELTTQQLRQROQSDAL 212
Qy 162 EQQOENQ-----AAINTVAANQETIAQNTNALNTQAAQ----- 195
Db 213 RGLQGEQTKLAE LRDRTRQADALAE LORSAGQQAQAVAVRTOQAALTAQTIDS LVGNV 272
Qy 196 -----LEAAQLNLQAE-----LTTAQDOKATLVA-- 219
Db 273 VERTFLEERRRRLEERREAEARRIREAQERARKEARLARIRAEQERKAREAAEA 332
Qy 220 --QKAAAEARQAAAAAQAABAKAAAEAKALQE-----QAAQAAAAANNNTQATDSDQQ 273
Db 333 ARQKALAE--ARAAQAKVAEAKARAEAKARAEQAARAAQAARAAQAARAAQARAQAE 390
Qy 274 AAAADNTQAAQTGDSQAQAQAVNNSDQESTTA-----TEAQPAS-- 315
Db 391 AAARAQAQAAAAAARAEQAARAAQQAQQAQAAQKVRQAVAREQDILQTOCQQAQOE 450
Qy 316 ---SASTAVVTANTSSANTYPAGQCTGWKSLAPWGNVGNWGQVAAASAAAAGYRVGST 372
Db 451 KQLAEALPPLAISRDGLFPPL---GKVAAP-----YGTSGAQEVLVSAGSGRAVAA 502
Qy 373 PSAGAVA-----VWMDGGYGHVAVYVTVGGGQIQV 402
Db 503 QTGNVIASVYAAAGWVILLDHGNSVITGYFGLQDTLVEV 542

RESULT 10
AH1387
cell wall binding proteins homolog lmo2504 [imported] - Listeria monocytogenes (strain E
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH1387
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Psihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-436 <GLA>
A:Cross-references: UNIPROT:Q8Y4E2; UNIPARC:UPI0000055221; GB:NC_003210; PIDN:CAD00582.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2504

Query Match 12.3%; Score 259; DB 2; Length 436;
Best Local Similarity 22.8%; Pred. No. 7.5e-06;
Matches 97; Conservative 90; Mismatches 152; Indels 86; Gaps 14;
Qy 6 LSAVLVSG-VTLSSATTLTSAV-----KADDFDAQIASODSKNNLTAAQQAQAV 55
Db 11 LSLIIISAPLTGAHADTINDMKRQNEIEQKSEIDKNIDSKNSLNHLESAEKDIAKEL 70
Qy 56 NTIQGV-----SALQTQQAELQAEORLEAQAATLQ---QIQTLSKIVARNESLKQQAARS 111
Db 71 ESLMKSLDNTKKLKEQEDKVSSNEKLA---KLQKEMEKLRNDRDRQKVLNRAI 126
Qy 112 QKSNATSYNAIINSGSVSDAINRVSAIREVVSANEKMLQOQ---EODKAAVEQKQOENQ 169
```

```
Db 127 QTTGTATSYLDIMIFRADDFKELVDRVTVVVSIVKADQNTMQDKQDDKLVASTSEKK 186
Qy 170 AA-INTVAANQETIAQNTNALNTQQAELQAEAAQLNLQAEELTTAQDQKATLVAQKAAAEAA 228
Db 187 LENLKVLAVELEVSXNNMESQKQKNDLVALAN-KKDLT--KSEQTLTLLASEQALTDDEE 243
Qy 229 RQAAAAQAAAAEAKAAEAKALQEQAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDS 288
Db 244 KRLASNIAGEKAKQEAIAKAEKRMQEAANA-----S 276
Qy 289 TQSQAAAAVNSDQSTTATEAQPASSA-----STAVVTANTSSANTYPAGQCTGW 341
Db 277 SAKSAAV---KQPSSSNEATETVSSCGGQFIRKPSGILTSGFSERTN----- 322
Qy 342 KSLAPWGNVWGN-----GGQMAASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTG 394
Db 323 ----PVTGKYESHKGDIAGGTVTVSAAASGTVVFSFGFASGSGF---GGYGVVVKIDH 375
Qy 395 VGGQ 399
Db 376 GNGFQ 380

RESULT 11
H84099
cell wall-binding protein BH3600 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: H84099
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: H84099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <STO>
A:Cross-references: UNIPROT:Q9KX4; UNIPARC:UPI00000C424C; GB:AP001519; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3600

Query Match 12.2%; Score 257.5; DB 2; Length 461;
Best Local Similarity 22.1%; Pred. No. 9.5e-06;
Matches 98; Conservative 75; Mismatches 180; Indels 91; Gaps 11;
Qy 31 DAQIASQDSKNNLTAAQQAQAAQVNTIQGVSAQTQQAELQAEORLEAQAATLQOOI 90
Db 56 EAELEKEVELGDTTAEIERLDKEVEETSGKQEKREEIEEVQAELEEK-----EQI 108
Qy 91 QTLSSKIVARNESLKQQAASQAQSNAAATSYINAIINSGSVSDAINRVSAIREVVSANEKM 150
Db 109 EILEERIAERDELLKDRARAWYQNGSGIDYLEVILGAKSFGDFLDRVSALSIVA----- 162
Qy 151 LQQQODKAAVQKQOENQAALNTVAANOETTAAQNTNALNTQQAELQAEAAQLNLQAEELTTA 210
Db 163 ----EQDRGILEAHIEDHRL-----EAKAQVEEKLETLEGLHLELE-NLMAQLEEQ 210
Qy 211 QDOKATLVAQAAAAEAAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 259
Db 211 QKEKEKVMGELASRDELHGLDLENDLELRQKEKALQEYELWKKQEBERKAAEKAA 270
Qy 260 ANNTQATDASDQAAAAADNTQAAQTGDSQAQAQAQAVNNSDQESTTATEAQPAS-SAS 318
Db 271 AEAQAQAQASGSGSGSSSDSGNSGTTSRNSGSSSGSGGGE-----TGSVPSSSGSGF 326
Qy 319 TAVVTANTSSANTYPAGQCTGWKSLAPWGNVGNWGQVAAASAAAAGYRVGSTPSAGAV 378
Db 327 MRPATGDISPPFGYRTHPVT-QQRKLHAGIDIRRGNRNSVNVVVAAYDGTWVQST----- 379
Qy 379 AVWMDGGYGHVAV-----TGVOGG-- 398
Db 380 --YSSGGYGNVILIAHSYNGRQVTLIAHLETRSVSAGORVSKGQTIGIMTGLTGPH 437
```


A:Gene: tola

A:Map position: 17 min

A:Start codon: GTG

C:Keywords: nucleotide binding; P-loop; transmembrane protein

F:14-34/Domain: transmembrane #status predicted <HSS>

F:78-301/Domain: helical #status predicted <HSS>

F:355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 11.0%; Score 230.5; DB 2; Length 421;
Best Local Similarity 28.2%; Pred. No. 0.00017;
Matches 116; Conservative 67; Mismatches 177; Indels 51; Gaps 19;

QY 5 ILISVLVSGVTSSATTLGAVKADDDPAQIASQ-----DSKINNLTAAQQAAQAQVNTIQ 59

DB 16 IISAVL--HVILFAALIWSS-----FDENIEASAGGGSSIDAVNVDSGAVVQYKRMQ 68

QY 60 GVISA-----LQTOQA--ELQ-----AEORL--EAQSATLGGQIQITLSSKIVARNES 103

DB 69 SQESSAKRSDEQRKMKKEQQAABELRQAAEQERLQLEKERLAAEQKKQAEAAKQAE 128

QY 104 LKQ-QARSQAQKNAATSYINAISKVSVDAINRVSAIREVVSANEKMLQQQEQDQKAAVE 162

DB 129 LKQKQAEAAAKAADAQAKAEADAKAEAAK-----AAADAKKAEAEAAKAAAE 181

QY 163 -QKQENQAAI--NTVAANOETIAQNTNALNTQQAQLEAAQLNLQAEI--TTAQDQKATLV 218

DB 182 AQKKAEEAAALKKKAEAAEAAAEAAAEARKKAATEAEAKAEKAAEKAAAEKAAADKKA--A 239

QY 219 AQKAAE--FAAQQA 274

DB 240 ABEKAAADKKAQA 299

QY 275 AAADNTQAAQTGDTSTQSAQAQVNNSDQSTTATEAQ-PSASSASTAVVTANTSSANTYP 333

DB 300 ELSSGKNAPKTKGGAGKNASPAGSGNTKNGSGADINNYAQIKSAIESKFYDASSYA 359

QY 334 AGCCTGWGKSLAP--WVGNYWNGGQWA-ASAAAAGYRVGSTPSAGAVVW 381

DB 360 GKTCTLRIR-LAPDGMILLDKPEGDPALCQAALAAAKLAKIPKPPSQAVY 409

RESULT 15

B90835 probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: B90835

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaewara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B90835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-971 <HAY>

A:Cross-references: UNIPROT:Q8XDD4; UNIPARC:UPI000016542B; GB:BA000007; PIDN:BA035073.1;

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1650

Query Match 10.8%; Score 227; DB 2; Length 971;
Best Local Similarity 23.8%; Pred. No. 0.00061;
Matches 105; Conservative 76; Mismatches 205; Indels 56; Gaps 10;

QY 19 ATTLSAVKADDDPAQIASQDSKINNLTAAQQAAQAQVNTIQGVSAIQTOQAELQAEOR 78

DB 121 AQNTAAKKSADASTSAREATHATDAADSAFAASTSAGQAASSAQSSAGTAGTATKA 180

QY 79 LEAQSATLGGQIQITLSSKIVARNESLQQAARSQAQKNAATSYINATINSKVSVDAINRV 138

DB 181 TEA-----SKSAAAEKSAATSAAGA-AKTSETNAVQSQA-----TS 221

Search completed: February 15, 2006, 18:10:25
Job time : 21.9676 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:21 ; Search time 123.476 Seconds
(without alignments)
2462.693 Million cell updates/sec

Title: US-10-797-821-33
Perfect score: 2105
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGNFNPQSVIYYPN 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2105	100.0	431	2 Q938V0_STRMU	Q938v0 streptococc
2	2096	99.6	431	2 Q9AG98_STRMU	Q9ag98 streptococc
3	2092	99.4	431	2 Q938V3_STRMU	Q938v3 streptococc
4	2092	99.4	431	2 Q8DMV3_STRMU	Q8dmv3 streptococc
5	2075.5	98.6	432	2 Q938V2_STRMU	Q938v2 streptococc
6	2066.5	98.2	432	2 Q938V1_STRMU	Q938v1 streptococc
7	1238	58.8	447	2 Q9AKA4_STRAG	Q9aka4 streptococc
8	1238	58.8	447	2 Q8E2H1_STRAS	Q8e2h1 streptococc
9	1238	58.8	447	2 Q8E7X9_STRAS	Q8e7x9 streptococc
10	1095	52.0	474	2 Q5M6K4_STRT2	Q5mek4 streptococc
11	1092.5	51.9	398	2 Q5XEL1_STRP6	Q5xel1 streptococc
12	1092.5	51.9	398	2 Q7CNQ7_STRP8	Q7cnq7 streptococc
13	1092.5	51.9	398	2 Q8P318_STRP3	Q8p318 streptococc
14	1087.5	51.7	485	2 Q5M212_STRT1	Q5m212 streptococc
15	1085.5	51.6	398	2 Q9A1Z8_STRPY	Q9a1z8 streptococc
16	1057.5	50.2	392	2 Q8DMV4_STRR6	Q8dmv4 streptococc
17	1057.5	50.2	392	2 Q97N55_STRPN	Q97n55 streptococc
18	947	45.0	211	2 QZAS7_STRMU	Qzasa7 streptococc
19	666	31.6	461	1 USF45_IACLC	P22865 lactococcus
20	661.5	31.4	456	2 Q9CDJ1_LACLA	Q9cdj1 lactococcus
21	598	28.4	524	2 Q9K2J9_ENTFC	Q9k2j9 enterococcu
22	583	27.7	516	1 P54_ENTFC	P13692 enterococcu
23	566.5	26.9	576	2 Q9KJJ3_ENTHR	Q9kjj3 enterococcu
24	549.5	26.1	482	2 Q5MSM6_STRK2	Q5msm6 streptococc
25	548.5	25.2	470	2 Q8QE3_9LACT	Q8rqe3 lactococcus
26	529.5	25.1	482	2 Q93LK4_ENTFA	Q93lk4 enterococcu
27	514	24.4	461	2 Q56SA7_STRTR	Q56sa7 streptococc
28	511	24.3	449	2 Q93LK3_ENTFA	Q93lk3 enterococcu
29	436	20.7	211	2 Q8DVU8_STRMU	Q8dvu8 streptococc
30	420.5	20.0	544	2 Q840X3_STRMU	Q840x3 streptococc
31	418.5	19.9	544	2 Q840V8_STRMU	Q840v8 streptococc

32	415.5	19.7	226	2 Q9JRG6_STRIT	Q9jrg6 streptococc
33	415.5	19.7	544	2 Q840W6_STRMU	Q840w6 streptococc
34	413.5	19.6	544	2 Q8DUR7_STRMU	Q8dur7 streptococc
35	399.5	19.0	129	2 Q5M137_STRT1	Q5m137 streptococc
36	399	19.0	169	2 Q8E3F4_STRAS	Q8e3f4 streptococc
37	392	18.6	169	2 Q8DXT4_STRAS	Q8dxt4 streptococc
38	359.5	17.1	398	2 Q927Y8_LISIN	Q927y8 listeria in
39	346	16.4	401	2 Q71WS3_LISMP	Q71ws3 listeria mo
40	346	16.4	401	2 Q9RE04_LISMO	Q9re04 listeria mo
41	335.5	15.9	604	2 Q6M552_CORGL	Q6m552 corynebacte
42	334	15.9	430	2 Q81HV2_BACCR	Q81hv2 bacillus ce
43	329.5	15.7	600	2 Q8NQA0_CORGL	Q8nqa0 corynebacte
44	326.5	15.5	427	2 Q73DGO_BACCI	Q73dgo bacillus ce
45	319.5	15.2	473	1 YVCE_BACSU	P40767 bacillus su

ALIGNMENTS

RESULT 1
Q938V0_STRMU
ID Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates";
RT Infect. Immun. 69:6931-6941 (2001).
RL EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8B8C4609F CRC64;

Query Match	100.0%;	Score 2105;	DB 2;	Length 431;
Best Local Similarity	100.0%;	Pred. No. 2.9e-86;	Mismatches 0;	Indels 0;
Matches 431;	Conservative 0;	Gaps 0;		
Qy	1	MKKRILSAVLVSGVTLSATLSAVKADFDQAIASQDSKINNLTAAQQAQAQVNTIQG	60	
Db	1	MKKRILSAVLVSGVTLSATLSAVKADFDQAIASQDSKINNLTAAQQAQAQVNTIQG	60	
Qy	61	QVSALOTQAELOAENRLEAQSGATLGGQIQTLSSKIVARNESLKQOAFSAQKSAATSY	120	
Db	61	QVSALOTQAELOAENRLEAQSGATLGGQIQTLSSKIVARNESLKQOAFSAQKSAATSY	120	
Qy	121	INAIINSKSVSDAINRVSAIREVVSANEKMLQQQDQDKAAVEQKQENQAATNTVAANO	180	
Db	121	INAIINSKSVSDAINRVSAIREVVSANEKMLQQQDQDKAAVEQKQENQAATNTVAANO	180	
Qy	181	TTAQNTNALNTQQAQLEAAQLNLQALTTAQDQKATLVAQKAAAEARQAQAAAAEA	240	

Db 181 TIAQNTNALNTQQAQLAEAAQLNLQALHTTAQDOKATLVAQKAAAEARQAAAAQAAEA 240
 QY 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAQVNN 300
 Db 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAQVNN 300
 QY 301 DOESTTATEAQPSSASASTAVTANTSSANTYPAGCTGKSLAPWVGNVWNGGQWAA 360
 Db 301 DOESTTATEAQPSSASASTAVTANTSSANTYPAGCTGKSLAPWVGNVWNGGQWAA 360
 QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGIIOVEANYAGNOSIGNYRGWF 420
 Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGIIOVEANYAGNOSIGNYRGWF 420
 QY 421 NPGSVSYIYPN 431
 Db 421 NPGSVSYIYPN 431

RESULT 2

Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
 AC Q9AG98;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)
 DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
 GN Name=sagA;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21153617; PubMed=11254612;
 RX DOI=10.1128/IAI.69.4.2493-2501.2001;
 RA Chia J.-S., Lee Y.-Y., Huang P.-T., Chen J.Y.;
 RA "Identification of stress-responsive genes in Streptococcus mutans by differential display reverse transcription-PCR."
 RT Infect. Immun. 69:2493-2501(2001).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21481977; PubMed=11598074;
 RX DOI=10.1128/IAI.69.11.6987-6998.2001;
 RA Chia J.-S., Chang L.-Y., Shun C.T., Chang Y.Y., Chen J.Y.;
 RA "A 60-kilodalton immunodominant glycoprotein is essential for cell wall integrity and the maintenance of cell shape in Streptococcus mutans."
 RT Infect. Immun. 69:6987-6998(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21481971; PubMed=11598068;
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 RA Duncan M.J.;
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding protein B and analysis of genetic diversity and protein production in clinical isolates."
 RT

RL Infect. Immun. 69:6931-6941(2001).
 DR EMBL; AF338445; AAK08104.1; -; Genomic DNA.
 DR EMBL; AY046411; AAK94501.1; -; Genomic DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5E47232E CRC64;
 Query Match 99.6%; Score 2096; DB 2; Length 431;
 Best Local Similarity 99.5%; Pred. No. 7,3e-86;
 Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQSKINNLTAAQAAAAQAVNTIQG 60
 Db 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQSKINNLTAAQAAAAQAVNTIQG 60
 QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
 QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEKQKQENQAATNTVAANO 180
 Db 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQEQDKAAVEKQKQENQAATNTVAANO 180
 QY 181 TIAQNTNALNTQQAQLAEAAQLNLQALHTTAQDOKATLVAQKAAAEARQAAAAQAAEA 240
 Db 181 TIAQNTNALNTQQAQLAEAAQLNLQALHTTAQDOKATLVAQKAAAEARQAAAAQAAEA 240
 QY 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAQVNN 300
 Db 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAQVNN 300
 QY 301 DOESTTATEAQPSSASASTAVTANTSSANTYPAGCTGKSLAPWVGNVWNGGQWAA 360
 Db 301 DOESTTATEAQPSSASASTAVTANTSSANTYPAGCTGKSLAPWVGNVWNGGQWAA 360
 QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGIIOVEANYAGNOSIGNYRGWF 420
 Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGIIOVEANYAGNOSIGNYRGWF 420
 QY 421 NPGSVSYIYPN 431
 Db 421 NPGSVSYIYPN 431
 RESULT 3
 Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
 AC Q938V3;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Glucan-binding protein B.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
 RA "Cloning of the gbpB gene from Streptococcus mutans."
 RT J. Dent. Res. 79:224-224(2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=21481971; PubMed=11598068;
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 RA Duncan M.J.;
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding protein B and analysis of genetic diversity and protein production in clinical isolates."
 RT


```
RT clinical isolates.";
RL Infect Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AEE50E9 CRC64;

Query Match      98.6%; Score 2075.5; DB 2; Length 432;
Best Local Similarity 98.6%; Pred. No. 5.9e-85;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAQQOKATLVAQKAAAEAAARQAAAAQA 240
Db 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAQQOKATLVAQKAAAEAAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 299
Db 241 KAAAEKALQEQAAQAAQAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 300
QY 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359
Db 301 SDQESTTATAAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 360 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANAGNOSIGNYRGW 419
Db 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANAGNOSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 6
Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gopB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RT Clinical isolates.";
RL Infect Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match      98.2%; Score 2066.5; DB 2; Length 432;
Best Local Similarity 98.4%; Pred. No. 1.5e-84;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAQQOKATLVAQKAAAEAAARQAAAAQA 240
Db 181 TTAQNTNALNTQQAQLEAAQNLQAEELTTAQQOKATLVAQKAAAEAAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 299
Db 241 KAAAEKALQEQAAQAAQAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 300
QY 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359
Db 301 SDQESTTATAAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 360 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANAGNOSIGNYRGW 419
Db 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANAGNOSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
DE Name=pcsB;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183(2001).
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```



```
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Signal.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.8%; Score 1238; DB 2; Length 447;
Best Local Similarity 60.7%; Pred. No. 1.1e-47;
Matches 280; Conservative 43; Mismatches 94; Indels 44; Gaps 10;

QY 1 MKKRILSAVLVSGVTLSAATLSSATLSSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSAATLSSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 58
QY 61 QVSALQTOQAELEAQNORLEAQSATLGGQIQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 59 QVGALESQOSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLKQVRSQAQGN-LTNY 117
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQODKAAVEKQOENQAANTVAANOE 180
DB 118 INTILNSKSVDAVNRVVAIREVVSANERKMLAQOQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TIAQNTNALNTQQAQLEAQAQLEAQLTAAQOKATLVAQKAAAEAAQAAQAAQAAEA 240
DB 178 AIENKKAALATQRAQLEAQAQLEAQLTAAQOKATLVAQKAAAEAAQAAQAAEA 237
QY 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQQAADNTQAAQTGDSSTE--QSAQAQVYN 298
DB 238 KAAAEKALQEQAAQAAQAAAE-----QVESATAPTETVQTPRTEIKPSNLTAIS 286
QY 299 NSDQSTTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336
DB 287 SATTVATTATATNEPKVTPQPSVTVKAEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMGQ 346
QY 337 CTWGVKSLAPVGNVYNGWGQWAAASAAAAGYRGVSTPSSAGAVAVM--NDGGYGHVAVYTG 394
DB 347 CTWGAASMASVGNVYNGWGNQWASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVVTS 406
QY 395 V-QGGQIQVQEAANYAGNOSIGNYRWFNP---GSVSYIYPN 431
DB 407 VANNSSIQWMSYAGNMSIGNYRGSFNPSPSAGSVYIYPN 447

RESULT 8
QBE2H1_STRAS5 PRELIMINARY; PRT; 447 AA.
AC QBE2H1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocustNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Khorov N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carly H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae.";
Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
EMBL; AE014192; AAM98925.1; -; Genomic_DNA.
TIGR; SAG0017; -.
InterPro; IPR007921; CHAP.
InterPro; IPR009148; Siba.
Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.8%; Score 1238; DB 2; Length 447;
Best Local Similarity 60.7%; Pred. No. 1.1e-47;
Matches 280; Conservative 43; Mismatches 94; Indels 44; Gaps 10;

QY 1 MKKRILSAVLVSGVTLSAATLSSATLSSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSAATLSSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 58
QY 61 QVSALQTOQAELEAQNORLEAQSATLGGQIQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 59 QVGALESQOSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLKQVRSQAQGN-LTNY 117
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQODKAAVEKQOENQAANTVAANOE 180
DB 118 INTILNSKSVDAVNRVVAIREVVSANERKMLAQOQADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TIAQNTNALNTQQAQLEAQAQLEAQLTAAQOKATLVAQKAAAEAAQAAQAAQAAEA 240
DB 178 AIENKKAALATQRAQLEAQAQLEAQLTAAQOKATLVAQKAAAEAAQAAQAAEA 237
QY 241 KAAAEKALQEQAAQAAQAAANNTQATDASDQQAADNTQAAQTGDSSTE--QSAQAQVYN 298
DB 238 KAAAEKALQEQAAQAAQAAAE-----QVESATAPTETVQTPRTEIKPSNLTAIS 286
QY 299 NSDQSTTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336
DB 287 SATTVATTATATNEPKVTPQPSVTVKAEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMGQ 346
QY 337 CTWGVKSLAPVGNVYNGWGQWAAASAAAAGYRGVSTPSSAGAVAVM--NDGGYGHVAVYTG 394
DB 347 CTWGAASMASVGNVYNGWGNQWASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVVTS 406
QY 395 V-QGGQIQVQEAANYAGNOSIGNYRWFNP---GSVSYIYPN 431
DB 407 VANNSSIQWMSYAGNMSIGNYRGSFNPSPSAGSVYIYPN 447

RESULT 9
QBE7X9_STRAS3 PRELIMINARY; PRT; 447 AA.
AC QBE7X9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocustNames=gs0016;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Ruzniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Ruzniok M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
invasive neonatal disease.";
Mol. Microbiol. 45:1499-1513 (2002).
EMBL; AL766843; CAD45661.1; -; Genomic_DNA.
```



```

RT M6 strain."
RL J. Infect. Dis. 190:727-738(2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match      51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 3e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VCAEDLSKIAKQDSIISNLTEQKAQNOVSALQA 57
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSQAQNSAATSY 120
DB 58 QVSSLSQSEODKLTARNTLEALSKRFEQIKALTSQIVARNEKLNQARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVSVANEKMLQQEQDKAAVEQKQOENQAINTVAANO 180
DB 118 INALLNSKISDVNRLVAINRAVSAKAKLLEQKADKVSLEEKQAANTAITNTIAAN 177
QY 181 TTAQNTNALNTQOALEAAQLNLOAELTTAODQKATLVAQKAAAEAEARQAAAAA 240
DB 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAQKEAAEKAAAEALAEQA 237
QY 241 KAAAEAKALQEQAAQAQAANNTQATDASDQQAADNTQAOTGDSQSAQAQVNN 300
DB 238 KAEQA---AQQAASVEAAKSAITPAQA-----TPAAQSSNAIEPAALTA--- 280
QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNNGGOWAA 360
DB 281 ----PAAPSARPQTS-----YDSSNTYPVQCCTGWGKSLAPWAGNNGGOWAY 326
QY 361 SAAAGYRGVSTPSAGAVAVNDGGYGHVAVYTGVO-GGQIQVOEANYAGNOSIGNY 419
DB 327 SQAAGYRGVSTPMVGAIAVNDGGYGHVAVVVEQSSASIRVMSNYSGRQYIADHR 386
QY 420 FNPQSVSYIYPN 431
DB 387 FNPTGVTFIYPH 398

RESULT 12
Q7CNQ7_STRP8
ID Q7CNQ7_STRP8 PRELIMINARY; PRT; 398 AA.
AC Q7CNQ7_STRP8
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=spvM18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=1917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).

DR EMBL; AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match      51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 3e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VCAEDLSKIAKQDSIISNLTEQKAQNOVSALQA 57
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSQAQNSAATSY 120
DB 58 QVSSLSQSEODKLTARNTLEALSKRFEQIKALTSQIVARNEKLNQARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVSVANEKMLQQEQDKAAVEQKQOENQAINTVAANO 180
DB 118 INALLNSKISDVNRLVAINRAVSAKAKLLEQKADKVSLEEKQAANTAITNTIAAN 177
QY 181 TTAQNTNALNTQOALEAAQLNLOAELTTAODQKATLVAQKAAAEAEARQAAAAA 240
DB 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAQKEAAEKAAAEALAEQA 237
QY 241 KAAAEAKALQEQAAQAQAANNTQATDASDQQAADNTQAOTGDSQSAQAQVNN 300
DB 238 KAEQA---AQQAASVEAAKSAITPAQA-----TPAAQSSNAIEPAALTA--- 280
QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNNGGOWAA 360
DB 281 ----PAAPSARPQTS-----YDSSNTYPVQCCTGWGKSLAPWAGNNGGOWAY 326
QY 361 SAAAGYRGVSTPSAGAVAVNDGGYGHVAVYTGVO-GGQIQVOEANYAGNOSIGNY 419
DB 327 SQAAGYRGVSTPMVGAIAVNDGGYGHVAVVVEQSSASIRVMSNYSGRQYIADHR 386
QY 420 FNPQSVSYIYPN 431
DB 387 FNPTGVTFIYPH 398

RESULT 13
Q8P318_STRP3
ID Q8P318_STRP3 PRELIMINARY; PRT; 398 AA.
AC Q8P318_STRP3
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=Spv0015, SpvM3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS315 / Serotype M3;

```

```
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Berses S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phase-encoded toxins, the high-virulence phenotype, and clone
RT emergence";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL; BA000034; BAC63110.1; -; Genomic DNA.
DR EMBL; AB014136; BAM78621.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 38-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLGATT---VGAEDLSTKIAKQDSIIISNLTTEKKAQNOVSALQA 57

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSQAOKSNAATSY 120
DB 58 QVSSLSQEQDKLTARNTLEALSKEPQEI KALTQIIVARNESKLNQARSYKNNETSGY 117

QY 121 INAINSKSVDAINRVSAREVSNANRKLQOQODKAAVEKQOENQAANTVAANOE 180
DB 118 INALLNSKISDVAVRLVAINRAVSAKALKEQQKADRVLSBEKQAANQTALNTAANNA 177

QY 181 TTAQNTNALNTQOALEAQLNQLAELTTAODKATLVAQKAAAEAEARQAQAAAEAA 240
DB 178 MAEENQNTLRTQANLEAATATLANLQASATEDKANLVAQKAAAEAAEAALQAQAKV 237

QY 241 KAAAEAKALQEQAAQAAANNNTQATDASDQQAADNDTQAAQTGDSQEQSAAQAVNNS 300
DB 238 KAQEQA---AQQAASVERAKSAITPAQ-----TPAQSNAIEPAALTA---- 280

QY 301 DQESTTATEAOPSASSASTAVTANTSSANTYPAGCTGWKSLAPWVGNVWNGGOWAA 360
DB 281 ----PAAPSARPOTS-----YDSSNTYPVGOCTGWAKSLAPWAGNNGGOWAY 326

QY 361 SAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGVO-GGQIQVQEAANYAGNQSIGNYRGW 419
DB 327 SAQAAGYRTGSTPMVGAIAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQVIADHRCW 386

QY 420 FNPGSVSYIYPN 431
DB 387 FNPGTGVTIYPH 398

RESULT 14
QSM212_STR11
ID Q5M212_STR11 PRELIMINARY; PRT; 485 AA.
AC Q5M212;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN Name=pcdB; OrderedLocusNames=str0022;
OS Streptococcus thermophilus (strain CNRZ 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=299768;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
```

```
RA Kulakauskas S., Lapidus A., Goltsman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcour J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV61641.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 51.7%; Score 1087.5; DB 2; Length 485;
Best Local Similarity 50.8%; Pred. No. 68-41;
Matches 248; Conservative 67; Mismatches 112; Indels 61; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYSQIAATNNAISNLASQOEAAQAQVATIOS 57

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQOARSQAOKSNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQARSQAQNTATSY 117

QY 121 INAINSKSVDAINRVSAREVSNANRKLQOQODKAAVEKQOENQAANTVAANOE 180
DB 118 INSILNSKISSEAITRITAISKVVTANNLLTKQESDQKELAAKQENQAANTTAANKS 177

QY 161 TTAQNTNALNTQOALEAQLNQLAELTTAODKATLVAQKAAAEAEARQAQAA---AQA 236
DB 178 ELETTEAGLTTCQAELEAAQVTLAAELATAQNEKTSLSVSAKSTASVASTASVAQSOA 237

QY 237 AAEAKAAAEAKALQEQA---AQQAANNNTQA---TDAQDQQA---AADNTQAAQTDG 287
DB 238 TAESNTAQVVASSEATSVASSEVAATSEVAQPSFVSEISTASEAQEPASSETSE 297

QY 288 STEQSAQAQVNN-----SDQESTTAT-----EAQPSASSASTAVVTANT 326
DB 298 VQPSAAPAVSEAPASVAPVATSEAAPATSEAPASVAPVATSEAPASVAPVAPVAPV 357

QY 327 SSA-----NTYPAGCTGWKSLAPWVGNVWNGGOWAAASAAAAGYRVGS 371
DB 358 SEAAPAAETHKVSAASTNTYPVGOCTGWKSLAPWAGNNGWNAKNWTASAAQAGHSVGT 417

QY 372 TPSAGAVAVW-ND-GGYGHVAVYTGVOG-GQIQVQEAANYAGNQSIGNYRGWFNP----- 422
DB 418 TPVAGAIAPVNDGGYGHVAVYVTSAGANSIQVMESYAGNWSISNYRGTFTDPTSSAHG 477

QY 423 GSVSYIYP 430
DB 478 GSVFYIYP 485

RESULT 15
Q9A128_STRPY
ID Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; O7BHS9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
```

```

RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF006474; AAK33158.1; -; Genomic DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PROSITE; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome; Signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;

Query Match 51.6%; Score 1085.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 6.1e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATTLSAVKADDFDAQIASODSKINNLTAAQOAAQAVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1 MKKRILSAVLVSGVTLGAATT--VGAEDLSTKIADQSDIISNLTTEQKAAQNVSAQQA 57
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 QVSALQTOQAEIQAEORLEAQSATLIGQOIOTLSSKIVARNESLKKQARSAAKSNAATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 58 QVSSILQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQARSAYKNNETSGY 117
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 INAIINSKSVDAINRVSAIREVWSANEKMLQQQSDKAAVEKQOENQAAINTVAANOE 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 118 INALLNSKSIISDVNRLVAINRAVSAVSAKLLLEQKADKVSLEEKQAANQTINTIAANMA 177
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 TTAQNTNALNTQOALEAQLNLAELTTAQDQKATLVAKAAAEAEAPQAAAQAAAEAA 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 178 MAEENQNTLRTQOANLVAAATNLAQLASATEDKANLVAKAEAEKAAAEALAQSOAAKV 237
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSFEQSAQAQAVNNS 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 238 KA-----QEQA-----QQAASVEAAKSALTTPAQATPAQSSNAI 273
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGCTWGVKSLAPWGVNNGNGQWAA 360
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 274 EPAALTA----PAAPSGP---QTSYDSSNTYPVGQCTWGAKS LAPWAGNNNGNGQWAW 326
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 361 SAAAGYRVGSTPSAGAVVNDGGYGVHYVTGQV-GGQIOVQENYAGNQSIGNYRGW 419
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 327 SAQAAGYRTGSTPMVGAIAVNDGGYGVHYVVVEVQSSAIRVMESNYSGROYIADHRGW 386
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 420 FNPQSVSYIYPN 431
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 387 FNPTGVTFIYPH 398
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 29.9537 Seconds
(without alignments)
1189.611 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRGWFMFGSVIYIPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/ECTUS_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfilesl.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1057.5	50.2	399	2	US-09-107-433-3230
2	1056.5	50.2	392	2	US-09-583-110-4374
3	671	31.9	461	1	US-08-186-222-2
4	594	28.2	525	2	US-09-107-532A-5095
5	562	26.7	210	2	US-09-222-938A-67
6	526.5	25.0	497	2	US-09-134-000C-5990
7	507	24.1	449	2	US-09-071-035-482
8	507	24.1	449	2	US-10-206-576-482
9	507	24.1	450	2	US-09-134-000C-5714
10	458	21.8	422	2	US-09-071-035-484
11	458	21.8	422	2	US-10-206-576-484
12	278	13.2	469	2	US-09-489-039A-13565
13	235	11.2	257	2	US-09-710-279-3244
14	235	11.2	264	2	US-09-134-001C-5035
15	235	11.2	267	2	US-09-134-001C-4539
16	230.5	11.0	2310	2	US-09-874-923-120
17	228.5	10.9	477	2	US-09-902-540-11649
18	225.5	10.7	1236	2	US-09-769-787-109
19	220	10.5	468	2	US-09-328-352-6321
20	218.5	10.4	610	2	US-09-336-447A-11
21	218.5	10.4	610	2	US-09-952-267B-11
22	215	10.2	149	2	US-09-710-279-1682
23	215	10.2	157	2	US-09-710-279-2870
24	215	10.2	270	2	US-09-134-001C-5441
25	210	10.0	1566	1	US-08-687-956A-23
26	209.5	10.0	440	2	US-08-302-756B-35
27	209	9.9	266	2	US-09-134-001C-5453

ALIGNMENTS

RESULT 1

US-09-107-433-3230
; Sequence 3230, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...399
; SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
Sequence 14908, A
Sequence 12307, A
Sequence 13861, A
Sequence 23, Appl
Sequence 23, Appl
Patent No. 5352450
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 32, Appl
Sequence 4, Appl
Sequence 10005, A
Sequence 44973, A
Sequence 7646, Ap
Sequence 7647, Ap

US-09-107-433-3230

Query Match 50.2%; Score 1057.5; DB 2; Length 399;
Best Local Similarity 50.6%; Pred. No. 9.4e-70;
Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSATTLSAVKADDFDAQIASQDSKINNLTAQQAQAAQAVNTTQ 60
DB 8 MKKRILASLLSTVWVSQVAVLTTAAHETTTDKIAAQDNKI SNLTAQQAQAAQAVNTTQ 67

QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQITLSSKIVARNESLKKQARSQAQNAATSY 120
DB 68 QVSAIQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQ 127

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAANTVAANO 180
DB 128 INTIVNSKSITEAISRVAAMSEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 187

QY 181 TTAQNTNALNTQOALEAQLNLQAEELTTAODOKATLVAOKAAAEAEARQAAAAQAAEA 240
DB 188 KLADDAQALTTKQAEKKAELSLAAEKATAEKESKASLLEQKAAAEAEARQAAAAQAAEA 246

QY 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
DB 247 ----EKRAQQQSVLASANTNLTAQ-----VQAVSES 274

QY 301 DOESTTATEAOPSASASTAVTANTSSANTYPAGCCTGWKSLAPWVGNVYNGGOWAA 360
DB 275 -----AAAPVRKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAOWAT 322

QY 361 SAAAAAGYRVGSPSAGAVAVWNDGGYGHVAVYTVQG-GQIQVQEBYVAGNOSIGNYRGW 419
DB 323 SAAAAAGFTGTPQVGAACWNDGGYGHVAVVAVESTTRIQVSESNYAGNRTIGNHRGW 382

QY 420 FNP-----GSVSYIY 429
DB 383 FNPFTTSEGFYIY 397

RESULT 2

US-09-583-110-4374
; Sequence 4374, Application US/09583110
; Patent No. 6699703

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A

; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 4374
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4374

Query Match 50.2%; Score 1056.5; DB 2; Length 392;
Best Local Similarity 50.6%; Pred. No. 1.1e-69;
Matches 220; Conservative 72; Mismatches 92; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSATTLSAVKADDFDAQIASQDSKINNLTAQQAQAAQAVNTTQ 60
DB 1 MKKRILASLLSTVWVSQVAVLTTAAHETTTDKIAAQDNKI SNLTAQQAQAAQAVNTTQ 60

QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQITLSSKIVARNESLKKQARSQAQNAATSY 120
DB 68 QVSAIQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQ 127

DB 61 QVSAIQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQSNLQAEQ 120

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAANTVAANO 180
DB 128 INTIVNSKSITEAISRVAAMSEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 180

QY 181 TTAQNTNALNTQOALEAQLNLQAEELTTAODOKATLVAOKAAAEAEARQAAAAQAAEA 240
DB 188 KLADDAQALTTKQAEKKAELSLAAEKATAEKESKASLLEQKAAAEAEARQAAAAQAAEA 239

QY 241 KAAAEAKALQEAQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
DB 247 ----EKRAQQQSVLASANTNLTAQ-----VQAVSES 267

QY 301 DOESTTATEAOPSASASTAVTANTSSANTYPAGCCTGWKSLAPWVGNVYNGGOWAA 360
DB 275 -----AAAPVRKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAOWAT 315

QY 361 SAAAAAGYRVGSPSAGAVAVWNDGGYGHVAVYTVQG-GQIQVQEBYVAGNOSIGNYRGW 419
DB 323 SAAAAAGFTGTPQVGAACWNDGGYGHVAVVAVESTTRIQVSESNYAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429
DB 376 FNPFTTSEGFYIY 390

RESULT 3

US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007

; GENERAL INFORMATION:
; APPLICANT: Suri, Bruno
; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-186-222-2

Query Match 31.9%; Score 671; DB 1; Length 461;

Best Local Similarity 35.2%; Pred. No. 2.5e-41;
Matches 169; Conservative 103; Mismatches 140; Indels 68; Gaps 13;

QY 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKKIISAILMTSVILSAAPLSGVYADT-NSDIAKQDATISSAQSAKAQAQVDSLOS 59

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB 60 KVDLSQOKQTSTKAQIAKTESERKALNAQIATLINESIKERTKTLEAQAARSQAQVNSATNY 119

QY 121 INAINLSKVSDAINRVSAIREVVSANEXMLQOQEQDKAAVEQKQOENQAANTVAANOE 180
DB 120 MDAAVNSKSLTDVIOKVTAIATVSSANKOMLEQEQEKELSQKSVTKVKNYNQFVLSQ 179

QY 181 TTAQNTNALNTQOAELEAQLNQLAELTTAAQDOKATLVAKAAAEAAEAQAQAAQAAEA 240
DB 180 SLDQSAQELTSQAELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQAKQAAYEA 239

QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQOAAAADNTQAAQTGDSTEQ-----SAAQ 295
DB 240 Q-----QKEAAQAQAATAAKAVEAATSSASASSAPQVSTSTDTNTTNSASGN 291

QY 296 AVNNSDOESTTAEQPSASSASTAVVTANTSS-----ANTYP 333
DB 292 SGNSSNSSSSSSSSSSSSSSSNAGNTNGTSTGNTGTTTGGSGINSPIGNPYA 351

QY 334 AGOCT---WG-----VKSLAPVGNVNGGOWAASAAAAG--YRVGSTPSAGAVAV 380
DB 352 VGGCTDYWQYFAAQGIYIRNMP-----GNGGQWASNGPAQGVLVHVGAP--GVIAS 403

QY 381 -----WNDGGYGHVAVYTVQVQ-GQIQOQAEANYAGNQSIGNVGWFNPGSVSIYPN 431
DB 404 SPSADFGVANGPYGHVAVKVSNSDGTITIKEGGY-GTTWGHGHER-TVSASGVTFPMEN 461

RESULT 4
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5095:

SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...525
SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095

Query Match 28.2%; Score 594; DB 2; Length 525;
Best Local Similarity 31.0%; Pred. No. 1.3e-35;
Matches 162; Conservative 92; Mismatches 163; Indels 106; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 2 VKKSLISAVMVGSMTLTAVASPIAAADDFDQIQOQDKIADLKNQQAQDAQSDIDALES 61

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNAAATSY 120
DB 62 QVSEINTQADLLAQDTLRQESAQLVKDIADLQERIEKREDTIQKAREAQVSVNTSSNY 121

QY 121 INAINLSKVSDAINRVSAIREVVSANEXMLQOQEQDKAAVEQKQOENQAANTVAANOE 180
DB 122 IDAVLNADSLADAIGRVQAMTMTVKANNLMEQKQDKKAVEDKKAENDAKLKEAENQA 181

QY 181 TTAQNTNALNTQOAELEAQLNQLAELTTAAQDOKATLVAKA-----AAEE 226
DB 182 ALESQKGDLLSQADNLVLTSLAAEQATAEADKKADLNQKAEAEQAARIREQORLAEQ 241

QY 227 AARQAAAAQAQAAEAKAAEAKALQEQAAQ-AQAANNNTQATDAS-----DQOAAAAADNTQ 281
DB 242 ARQQAQAEKAEKAREQAQEAQATQASSTAQSSATESSATQSSWTESSSATQSSATE 301

QY 282 AAQTGDSTEQSAAQAVNNSDQSTTATEA-----OPSASSASTAVTANTSSANT 331
DB 302 ESTTPESSTESTAPESSESTTAPESSESTTTPESSESTTTPESSESTTTPESSTTEEST 361

QY 332 YPA-----GQCTGWGVKSLAP----- 346
DB 362 TPAPTTPSTDQSDVTGNGT-GSSTPAPTPTTPEQPKVTPAPAPSGSVNGAAIYAEAYK 420

QY 347 WYGN--YMGNG-----GOWAASAAAAGVYRVG-STPSAGAVAV 380
DB 421 YIGTFYVWGGKDPGDFDCSGFTRYVYVMQVTGRDIGGTVPQESAGTKISVSQAKAGDLLF 480

QY 381 W-NDGGYGHVAVYTVQVGGQIQOQAEANYAGNQSIGNVGWFNPGSVSIYPN 422
DB 481 WGSQGGTYHVAIALG--GGQ-YIHAPQPGESVKVGSVQ-WFAP 519

RESULT 5
US-09-222-938A-67
; Sequence 67, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT

ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67
Query Match 26.7%; Score 562; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 9e-34;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTTQ 60
DB 1 MKKKILASLLSTVMVSVQAVVTTAAETTDKIAAQNKNISLNTAAQQAQAQVNTTQ 60
QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIOTLSSKIVARNESLKQARSQAQSNAAATSY 120
DB 61 QVSALQAEQSNLOAENDRLQAESKLEGEITELSKNIVSRNQSLEKQARSQAQTNVATSY 120
QY 121 INAINSKSVDAINRVSIAIRVVSANEKMLQOQODKAAVFOKQOENQAANTVAANQE 180
DB 121 INTIVNSKSITEAISRVAAEMSEIVSANNKMLEQKQADKKAISEKQVANNDAINTVIANQ 180
QY 181 TTAQNTNALNTQQAQLAENLQALNLAELTTA 210
DB 181 KLADDAQALTTTKQAEKAAELSLAAEKATSY 210

RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134, 000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055, 778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

Query Match 25.0%; Score 526.5; DB 2; Length 497;
Best Local Similarity 31.0%; Pred. No. 1.1e-30;
Matches 150; Conservative 81; Mismatches 178; Indels 75; Gaps 11;
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTTQ 60
DB 16 LKGSVLALMVCSITLTSVALPSAFADEYDTKIQQQDOKINALTSQMSDAEAKVALEN 75
QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIOTLSSKIVARNESLKQARSQAQSNAAATSY 120
DB 76 DMVETAKQIDTITAKKNKLSSEVSKLYSISDLNVRIQKREYQMTKQARDVQVNGSDSI 135
QY 121 INAINSKSVDAINRVSIAIRVVSANEKMLQOQODKAAVFOKQOENQAANTVAANQE 180
DB 136 IDAVIDADSDAIDGRVQAVSTWMSANNELLEQKEDKATVETKTKNVEKQAELEAATK 195
QY 181 TTAQNTNALNTQQAQLAENLQALNLAELTTA 210
DB 196 ELNDKTESLTKUIQOEVANNDLEQRSEEQKGGFKQKKEAEKRLAEQARQAAK 255
QY 235 -----QAAAEAKAAAKALQEAQAQAQAANNNTQATDASDQQAADANTQAQTDGSTE 290
DB 256 KAEQQA 307
QY 291 QSAQAQVNNSDQES-----TTATEAOPS-----ASSASTA 320
DB 308 STTQETTTSTSTETESVVTTPVAAAEPEKEKEVPVTNPTTPEKNEAKPGNGVGTGKQA 367

QY 321 VVTANTSSANTYPA-----GOCTMGVKSLAPWVGNVWNGGQWAAASAAAAAGYRVG-STPS 374
DB 368 INAALADVNSVATCWNQPGECGLVSVRWLAAGGNGFGVGGNSGVVASGATQVSKSNVQ 427
QY 375 AGAVAVNDG-----GYGHVAVYTVGGGQGOIQVOEANYAGNQSIGNVRGWFNPGSVSY 427
DB 428 PGDVVQVESAYSPPDSWIGGVHTLVTVGSGGSGVQIVEAN-----NPGGSGY 473
QY 428 IYPN 431
DB 474 VSSN 477
RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-482

Query Match 24.1%; Score 507; DB 2; Length 449;
Best Local Similarity 31.6%; Pred. No. 2.6e-29;
Matches 142; Conservative 88; Mismatches 175; Indels 44; Gaps 12;
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTTQ 60
DB 1 VKKRLFASVLLCSLTLSAATPSIALADNVDKKIEEKQEISSLKAKQGLASQVSSLEA 60
QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIOTLSSKIVARNESLKQARSQAQSNAAATSY 120
DB 61 EVSSVFDESMLREKQKTLKAKSEQLQOEITLNLQRIEKNRNEAIKNQARDVQVNGQSTTM 120
QY 121 INAINSKSVDAINRVSIAIRVVSANEKMLQOQODKAAVFOKQOENQAANTVAANQE 180
DB 121 LDVLDADSVADAIISRQAVSTIVSANNNDLMQQKEDKQAVVDKKAENEKKVKQLEATA 180
QY 181 TTAQNTNALNTQQAQLAENLQALNLAELTTA 210

Db 242 KAKQAAKPAKAEVKAEPVASSSTTEAQAAPASSSATESSTQQTTTTPSTDNSATEN 301
QY 285 TGDSTEQSAQAV-----NNSDQSTTATEAOPSASSASTAVVTANTSSANTYPAGQCT 338
Db 302 TSSSSEQVPQPTPSDNGNGGQTGGGTVTPTPEPTPAPSADPTINALNLVLRQSLG--- 358
QY 339 WGVKSLAPVGNVWNGGQWAAASAAAGYRVGSTPSAGAVAVWNGGYGHVAVTVCGQG 398
Db 359 -----LRPVV-----WDAGLAASATARAQVEAGGIPNDH-----WSRG--DEVIAIMWAPGN 404
QY 399 QIQV---QEAANYAGNOSIGNYRGW-FNPG 423
Db 405 SVIMAWYNETNMVMTASGSG-HRDWEINPG 432
RESULT 10
US-09-071-035-484
; Sequence 484, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 484:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-484
Query Match 21.8%; Score 458; DB 2; Length 422;
Best Local Similarity 31.0%; Pred. No. 9.4e-26;
Matches 131; Conservative 81; Mismatches 166; Indels 44; Gaps 12;
QY 28 DFDQAQIASQDSKINNLTAQQAQAQVNTIQGVSAQTQQAELQAEQNRLQEAQATLG 87
Db 1 DNVDKKIEKNQESLSLAKQKDLASQVSSLEAEVSVFDESMALREKQKTLKAKSEQIQ 60
QY 88 QIQTLSSKIVARNELKQQAESAKSNATSYINAINSKVSDAINRVSAIREVVSAN 147
Db 61 QEITNLNRIEKRNEAKNQARDVQNGQSTTMDLADLDADSVADRAISRQAVSTIVSAN 120
QY 148 ERMLOQQEQDKAAVEKQKQENQAANTVAANOETIAONTNALNTQQAQLQAEAL 207

Db 121 NDLMOQQKEDQAVVDVKKAEKVKVQLQLEATEAELETKRQDLLLSKOSELNVNKKASLALQ 180
QY 208 TTAQDKATLVAQKAAAE-EAARQAAAAQAAAE-AKAAAEAKALQEQAAQAAAAANNNTQ 265
Db 181 SSAESSKAGLEKQKAAAEQARLAAEQKAAAEKAKQAAAKPAKAEVKAEPVASSSTTE 240
QY 266 -----ATDASDQAAA-----AADNTQAAQTGDSSTEQSAQAQAV-----NNSDQEST 305
Db 241 QAAPASSSATESSTQQTTTTPSTDNSATENTGSSSEQVPQPTPSDNGNGGQTGG 300
QY 306 TATEAOPSASSASTAVVTANTSSANTYPAGQCTGWKSLAPVWVGNVWNGGQWAAASAAA 365
Db 301 GTVTPTPEPTPAPSADPTINALNLVLRQSLG-----LRPVV-----WDAGLAASATARA 349
QY 366 GYRVGSTPSAGAVAVWNGGYGHVAVTVCGGQIQV---QEAANYAGNOSIGNYRGW-PN 421
Db 350 QVEAGGIPNDH-----WSRG--DEVIAIMWAPGNVIMAWYNETNMVMTASGSG-HRDWEIN 402
QY 422 PG 423
Db 403 PG 404
RESULT 11
US-10-206-576-484
; Sequence 484, Application US/10206576
; Patent No. 6913907
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369P1D1
; INFORMATION FOR SEQ ID NO: 484:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 484:
US-10-206-576-484
Query Match 21.8%; Score 458; DB 2; Length 422;
Best Local Similarity 31.0%; Pred. No. 9.4e-26;
Matches 131; Conservative 81; Mismatches 166; Indels 44; Gaps 12;

Qy	28	DDFDIAIQASQSKINNLTAQOQAAQAVNTIOGVSAQTQOAELOAENQRLAQASATLG	87
Db	1	DNVDKKIEEKQEISSLKAKQGDLASQVSSLEAEVSSVFDESMALREQKOTLKAKSEQLQ	60
Qy	88	QOIOTLSSKIVARNESLKOQARSQAOKSNAATSYINAIINSKVSDAINRVSAIREVVSAN	147
Db	61	QEIITNLNRIEKENBAIKNOARDVQVNGOSITMLDAVLDAVDSDAISRVQAVSTIVSAN	120
Qy	148	EKMLQQQEQDKAAPVQKQOENQAAINTVAANOQETIAQNTNALNTQOAELEAAQLNLOAEL	207
Db	121	NDLMOQQQEDKQAVVDYKKAENEKVKYQLEATEAELETTRQDILLKQSELNVMKASLAEQ	180
Qy	208	TTAQDQKATLVAQKAAAB-EEARQAAAAQAAA-AKAAAAEKALQOAAQAAAAANNNTQ	265
Db	181	SSAESSKAGLEKQAAABAEQRLARBEQAAAEKAKQAAAKPAKEVKAEPVASSSTTE	240
Qy	266	-----ATDASDQAAA-----AADNTQAACTGDSTEQSAAQAV-----NNSDQEST	305
Db	241	AQAPASSSSSATESSTQQTETTTPTDNSATENTGSSSESQPVQPTTPSDNGNNGQTCG	300
Qy	306	TATEAQPSASASTAVTANTSSANTYPAGQCTGWGKSLAPWGVNYWNGGQWAAASAAA	365
Db	301	GTVTPTPTPTPAPSADPTINALNLVLRQSLG-----LRPVV-----WDAGLAASATARAA	349
Qy	366	GYYRGVSTESAGAVAVWNDGGYGHVAVYTGVOGQIQV---QEANYAGNQSIGNYRGW-FN	421
Db	350	QVEAGGITNDH-----WSRG--DEVIAIMWAPGNSVIMAWNYETNMVTSAGSG-HRDWEIN	402
Qy	422	PG	423
Db	403	PG	404

```

RESULT 12
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

```

[illegible]

Qy	255	QAQAAANNNTAATDASDOOAAAADNTOAAQTGDSTE-----QSA
Dd	310	AAREKAAD--KAKAKAAAAAKAAAKAAKENDGVNLLGDLSSGNAPKTGGGAAGNNNA
Qy	294	AOAVNDSQEETTATEAOPSASSASTAVVTANTSSANTYPACQCTGWGVKS LAP--WVGNY
Dd	368	AAAGSGMTKNSAGADINNYAGQTKSA-IESKFYDASSYAGTKTLRIK-LAPDGLLNI
Qy	352	WGNGGOWA-AASAAAAGYRVGSTPSAGAVAW 381
Dd	426	OSEGDDPALCOALAAAAAQAKFPKPPOAVY 456

```

RESULT 13
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent NO. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

```

	Query Match	11.2%	Score 235;	DB 2;	Length 257;
	Best Local Similarity	34.5%;	Pred. No. 1.1e-09;		
	Matches 59;	Conservative 29;	Mismatches 75;	Indels 8;	Gaps 5;
Qy	261	NNNTQATDASDQQAADNTQAACTGDTSTQCSAAQAVNNSDQESTTATEAQPSSASSASTA	320		
Db	91	NNYSNNYNNYQ---SNNTQSORKTQPTGGLGASYSTSSNNVHTTISA-PSNGVSL	146		
Qy	321	VVTANTSSANTYPAGQCTWGV-KSLAPWGVNYWNGQWAAASAAAAGYRGVSTPSAGAVA	379		
Db	147	--NARSAGNLVYSGQCTYYVDFRVGCKIGSTWGNANNWANAARSCYTVNNSPAKCAIL	204		
Qy	380	VWMDGGYGHVAYVTGV--QGQOIQVQEANYAGNOSIGNYRGFWNPNGSVSYIY	429		
Db	205	OTSOGAYGHVAYVEGVNNSGSIKRVSENNYGHGAGVVTSTRTISASQAQASVNY255			

```

RESULT 14
US-09-134-001C-5035
; Sequence 5035, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5035
; LENGTH: 264
; TYPE: PRT

```

```
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match      11.2%; Score 235; DB 2; Length 264;
Best Local Similarity 34.5%; Pred. No. 1.2e-09;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

QY 261 NNNTQATDASDQQAADNTQAAQTGDSFTEQSAQAQVNNSDQESTTATEAQPSSASASTA 320
DB 98 NNYYNSYNNYQ---SNTQSQRTTQPTGGLGASYSTSSNNVHTTSA-PSSNGVSL 153

QY 321 VVTANTSSANTYPAGQCTGTV-KSLAPWVGNVWNGGOWAASAAAAGYRVGSTPSAGAVA 379
DB 154 --NARSASGNLYTSQCCTYYFVDRVGGKIGSTWGNANNWANAARSGYTVNNSPAKGAIL 211

QY 380 VVNDGGYGHVAVTVGV-QGGQIQOVBANYAGNQSIGNVRGWNPGSVSYIY 429
DB 212 QTSQAYGHVAYVEGVNNGSIRVSEMYGHGAGVVTSTISASQAASNY 262

RESULT 15
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539
```

```
Query Match      11.2%; Score 235; DB 2; Length 267;
Best Local Similarity 34.5%; Pred. No. 1.2e-09;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

QY 261 NNNTQATDASDQQAADNTQAAQTGDSFTEQSAQAQVNNSDQESTTATEAQPSSASASTA 320
DB 101 NNYYNSYNNYQ---SNTQSQRTTQPTGGLGASYSTSSNNVHTTSA-PSSNGVSL 156

QY 321 VVTANTSSANTYPAGQCTGTV-KSLAPWVGNVWNGGOWAASAAAAGYRVGSTPSAGAVA 379
DB 157 --NARSASGNLYTSQCCTYYFVDRVGGKIGSTWGNANNWANAARSGYTVNNSPAKGAIL 214

QY 380 VVNDGGYGHVAVTVGV-QGGQIQOVBANYAGNQSIGNVRGWNPGSVSYIY 429
DB 215 QTSQAYGHVAYVEGVNNGSIRVSEMYGHGAGVVTSTISASQAASNY 265
```

Search completed: February 15, 2006, 18:11:59
Job time : 30.9537 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 91.8579 Seconds
(without alignments)
1960.467 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFMFGSVYIYPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:*

- 1: /cgn2_6/protdata1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/protdata1/pubpaa/US08_PUBCOMB.pep.*
- 3: /cgn2_6/protdata1/pubpaa/US09_PUBCOMB.pep.*
- 4: /cgn2_6/protdata1/pubpaa/US10A_PUBCOMB.pep.*
- 5: /cgn2_6/protdata1/pubpaa/US10B_PUBCOMB.pep.*
- 6: /cgn2_6/protdata1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2105	100.0	431	US-10-383-930-33	Sequence 33, Appl
2	2105	100.0	431	US-10-797-821-33	Sequence 33, Appl
3	2096	99.6	431	US-10-383-930-33	Sequence 30, Appl
4	2096	99.6	431	US-10-797-821-30	Sequence 30, Appl
5	2092	99.4	431	US-10-383-930-29	Sequence 29, Appl
6	2092	99.4	431	US-10-797-821-29	Sequence 29, Appl
7	2075.5	98.6	432	US-10-383-930-31	Sequence 31, Appl
8	2075.5	98.6	432	US-10-797-821-31	Sequence 31, Appl
9	2066.5	98.2	432	US-10-383-930-32	Sequence 32, Appl
10	2066.5	98.2	432	US-10-797-821-32	Sequence 32, Appl
11	1085.5	51.6	398	US-10-474-792-600	Sequence 600, App
12	1057.5	50.2	392	US-10-472-928-4652	Sequence 4652, Ap
13	1057.5	50.2	399	US-10-617-320-320	Sequence 3230, Ap
14	598	28.4	524	US-10-282-122A-57658	Sequence 57658, A
15	562	26.7	210	US-10-154-251-67	Sequence 67, Appl
16	507	24.1	449	US-09-071-035-482	Sequence 482, App
17	507	24.1	449	US-10-206-576-482	Sequence 482, App
18	507	24.1	449	US-10-912-362-482	Sequence 482, App
19	458	21.8	422	US-09-071-035-484	Sequence 484, App
20	458	21.8	422	US-10-206-576-484	Sequence 484, App
21	458	21.8	422	US-10-912-362-484	Sequence 484, App
22	335.5	15.9	630	US-10-494-674-6	Sequence 6, Appl1
23	329.5	15.7	600	US-09-738-626-5197	Sequence 5197, Ap
24	301	14.3	609	US-10-501-282-3184	Sequence 3184, Ap
25	300	14.3	440	US-10-501-282-3056	Sequence 3056, Ap
26	297.5	14.1	422	US-10-501-282-3054	Sequence 3054, Ap
27	288.5	13.7	377	US-10-501-282-3052	Sequence 3052, Ap

ALIGNMENTS

RESULT 1

US-10-383-930-33

; Sequence 33, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 33

; LENGTH: 431

; TYPE: PRT

; ORGANISM: Streptococcus mutans

; US-10-383-930-33

Query Match	Best Local Similarity	100.0%;	Score 2105;	DB 4;	Length 431;
Matches 431;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	MKKRILSAVLVSGVTLSATTL	SAVKADDFDAQIASQDSKINNLT	TAQQAQAQVNTIQG	60
Db	1	MKKRILSAVLVSGVTLSATTL	SAVKADDFDAQIASQDSKINNLT	TAQQAQAQVNTIQG	60
Qy	61	QVSALQTOQAELOAENORLEA	QSGATLGGQIQTLSSKI	VARNESLKQARS	120
Db	61	QVSALQTOQAELOAENORLEA	QSGATLGGQIQTLSSKI	VARNESLKQARS	120
Qy	121	INAINTSKVSDAINRVSAIRE	VSANEKMLQQQODKAAV	EOKQENQA	180
Db	121	INAINTSKVSDAINRVSAIRE	VSANEKMLQQQODKAAV	EOKQENQA	180
Qy	181	TTIAQNTNLTQQAQLEAAQL	NLQALTTAOPDKATL	VAQKAAAEFA	240
Db	181	TTIAQNTNLTQQAQLEAAQL	NLQALTTAOPDKATL	VAQKAAAEFA	240
Qy	241	KAAAEKALQEQAQAQA	AAAAAANNNTQATD	ASDOQAAAAA	300
Db	241	KAAAEKALQEQAQAQA	AAAAAANNNTQATD	ASDOQAAAAA	300
Qy	301	DOESTTATEAOPSASSA	STAVVTANTSSANTY	PAGCTWGVKSL	360
Db	301	DOESTTATEAOPSASSA	STAVVTANTSSANTY	PAGCTWGVKSL	360

QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 2
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1993-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match 100.0%; Score 2105; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 2e-120;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVQKQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAQLNLAELTTAODQKATLVAQKAAAEAAARQAAAAQAAAEA 240
DB 181 TTAQNTNALNTQOAELEAQLNLAELTTAODQKATLVAQKAAAEAAARQAAAAQAAAEA 240
QY 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
DB 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DQESTTATEAQPSSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNWYWGNGQWAA 360
DB 301 DQESTTATEAQPSSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNWYWGNGQWAA 360
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431
RESULT 3
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-30

Query Match 99.6%; Score 2096; DB 4; Length 431;
Best Local Similarity 99.5%; Pred. No. 7.1e-120;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSAATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVQKQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOAELEAQLNLAELTTAODQKATLVAQKAAAEAAARQAAAAQAAAEA 240
DB 181 TTAQNTNALNTQOAELEAQLNLAELTTAODQKATLVAQKAAAEAAARQAAAAQAAAEA 240
QY 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
DB 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DQESTTATEAQPSSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNWYWGNGQWAA 360
DB 301 DQESTTATEAQPSSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNWYWGNGQWAA 360
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTVGGGQIOVQEANYAGNOSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 4
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens


```
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match          99.6%; Score 2096; DB 5; Length 431;
Best Local Similarity 99.5%; Pred. No. 7.1e-120;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKVSDAINRVSAREVWSANEKMLQOQODKAAVEQKQOENQAANTVAANOE 180
DB 121 INAIINSKVSDAINRVSAREVWSANEKMLQOQODKAAVEQKQOENQAANTVAANOE 180

QY 181 TTAQNTNALNTOQAQLEAAQLNLQAEELTTAQPDKATLVAQKAAAEAAAFQAAAAQAABA 240
DB 181 TTAQNTNALNTOQAQLEAAQLNLQAEELTTAQPDKATLVAQKAAAEAAAFQAAAAQAABA 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300

QY 301 DOESTTATEAQPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360
DB 301 DOESTTATEAQPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360

QY 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
DB 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 5
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29

Query Match          99.4%; Score 2092; DB 4; Length 431;
Best Local Similarity 99.3%; Pred. No. 1.3e-119;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKVSDAINRVSAREVWSANEKMLQOQODKAAVEQKQOENQAANTVAANOE 180
DB 121 INAIINSKVSDAINRVSAREVWSANEKMLQOQODKAAVEQKQOENQAANTVAANOE 180

QY 181 TTAQNTNALNTOQAQLEAAQLNLQAEELTTAQPDKATLVAQKAAAEAAAFQAAAAQAABA 240
DB 181 TTAQNTNALNTOQAQLEAAQLNLQAEELTTAQPDKATLVAQKAAAEAAAFQAAAAQAABA 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300

QY 301 DOESTTATEAQPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360
DB 301 DOESTTATEAQPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360

QY 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
DB 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 6
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
```

```

; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match      99.4%; Score 2092; DB 5; Length 431;
Best Local Similarity 99.3%; Pred. No. 1.3e-119;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQOENQAALNTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQOENQAALNTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAOPDKATLVAQKAAABEAAARQAAAAQAABEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAOPDKATLVAQKAAABEAAARQAAAAQAABEA 240

QY 241 KAAAEAKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299

QY 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359
Db 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359

QY 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYTVGQGGQIQVQEAANYAGNOSIGNYRGW 419
Db 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYTVGQGGQIQVQEAANYAGNOSIGNYRGW 419

QY 420 FNPGSVSYIYPN 431
Db 420 FNPGSVSYIYPN 431
```

```

RESULT 7
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

Query Match      98.6%; Score 2075.5; DB 4; Length 432;
Best Local Similarity 98.6%; Pred. No. 1.3e-118;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
```

```

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQOENQAALNTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQOENQAALNTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAOPDKATLVAQKAAABEAAARQAAAAQAABEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAOPDKATLVAQKAAABEAAARQAAAAQAABEA 240

QY 241 KAAAEAKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299

QY 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359
Db 300 SDQESTTATEAQPSSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 359

QY 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYTVGQGGQIQVQEAANYAGNOSIGNYRGW 419
Db 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYTVGQGGQIQVQEAANYAGNOSIGNYRGW 419

QY 420 FNPGSVSYIYPN 431
Db 420 FNPGSVSYIYPN 431
```

```

RESULT 8
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR FILING DATE: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31
```

```

Query Match      98.6%; Score 2075.5; DB 5; Length 432;
Best Local Similarity 98.6%; Pred. No. 1.3e-118;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIQ 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQOENQAALNTVAANQE 180
```

```
Db 121 INAIINSKSVDAINRVSAIREVVSANEXKLOQOEQDKAAVEQKQENQAAINTVAANQE 180
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Qy 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 299
Db 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 300
Qy 300 SQESTTATEAOPSPASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db 301 SQESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Qy 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 420
Qy 420 FNPGSVSIYPN 431
Db 421 FNPGSVSIYPN 432

RESULT 9
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match 98.2%; Score 2066.5; DB 4; Length 432;
Best Local Similarity 98.4%; Pred. No. 4.5e-118;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Qy 61 QVSALQTQAELEAENQRLQEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALQTQAELEAENQRLQEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Qy 121 INAIINSKSVDAINRVSAIREVVSANEXKLOQOEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKSVDAINRVSAIREVVSANEXKLOQOEQDKAAVEQKQENQAAINTVAANQE 180
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Qy 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 299
Db 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 300
Qy 300 SQESTTATEAOPSPASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 300
Qy 300 SQESTTATEAOPSPASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
```

```
Db 301 SQESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Qy 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 420
Qy 420 FNPGSVSIYPN 431
Db 421 FNPGSVSIYPN 432

RESULT 10
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match 98.2%; Score 2066.5; DB 5; Length 432;
Best Local Similarity 98.4%; Pred. No. 4.5e-118;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Qy 61 QVSALQTQAELEAENQRLQEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALQTQAELEAENQRLQEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Qy 121 INAIINSKSVDAINRVSAIREVVSANEXKLOQOEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKSVDAINRVSAIREVVSANEXKLOQOEQDKAAVEQKQENQAAINTVAANQE 180
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQDQKATLVAKAAAEBAARQAAAAQAAAAA 240
Qy 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 299
Db 241 KAAAEAKALQEAQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGSDTQSAQAQAVNN 300
Qy 300 SQESTTATEAOPSPASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db 301 SQESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Qy 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVQGGQIQVQEANYAGNOSIGNYRGW 420
```

```
; 420 FNPQSVSYIYPN 431
; |||||
Db 421 FNPQSVSYIYPN 432

RESULT 11
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 600
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-600

Query Match 51.6%; Score 1085.5; DB 5; Length 398;
Best Local Similarity 53.7%; Pred. No. 2.6e-58;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAIASQDSKINNLTAAQQAQAQVNTTQ 60
DQ |||||
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAIASQDSKINNLTAAQQAQAQVNTTQ 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DQ |||||
Db 58 QVSSLOQSDQDKLTARNTLEALSUKRFEQEI KALTSQIVARNESLKQQAARSQAQSNAAATSY 117
QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQOQDQKAAVEQKQOENQAATINTVAANQE 180
DQ |||||
Db 118 INALLNKSISDVNRLVAINRAVSANAKLLSQKADKVSLEKQAAQOATINTIAANNA 177
QY 181 TIAQNTNALNTQOALEAAQLNQLAELTTAQQKATLVAQKAAAEAEARQAAAAQAAEA 240
DQ |||||
Db 178 MAEENQNTLRTOQANLVAATANLALQASATDKANLVAQKAAAEAEARQAAAAQAAEA 237
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNS 300
DQ |||||
Db 238 KA-----QEQAA-----QQAASVEAKSAITPAQATPAQAQSSNAI 273
QY 301 DQESTTATEAOPSASSASTAVVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWAA 360
DQ |||||
Db 274 EPAALTA---PAAPSAQF---QTSVDSNTYPVGCCTGWGAKSLAPWAGNWNWNGGQWAA 326
QY 361 SAAAGYRVGTPSPAGAVAVNDGGYGHVAVVTGQV-GQIQVQVANTAGNOSIGNYRGW 419
DQ |||||
Db 327 SAQAAGYRTGTPPMVGAIAVNDGGYGHVAVVVEVQSSIRWESNYSGROYIADHRGW 386
QY 420 FNPQSVSYIYPN 431
DQ |||||
Db 387 FNPQSVSYIYPN 398

RESULT 12
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W
```

```
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.E+01)
US-10-472-928-4652

Query Match 50.2%; Score 1057.5; DB 5; Length 392;
Best Local Similarity 50.6%; Pred. No. 1.3e-56;
Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAIASQDSKINNLTAAQQAQAQVNTTQ 60
DQ |||||
Db 1 MKKRILASLLSLTMVMSQVAVLTTHAETTTDDKIAAQNKKISNLTAAQQAQAQVNTTQ 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DQ |||||
Db 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQOQDQKAAVEQKQOENQAATINTVAANQE 180
DQ |||||
Db 121 INTIVNSKITEAISRAVMSSEIVSANKMLQOQDQKAAVEQKQOENQAATINTVAANQE 180
QY 181 TIAQNTNALNTQOALEAAQLNQLAELTTAQQKATLVAQKAAAEAEARQAAAAQAAEA 240
DQ |||||
Db 181 KLADDAQALTTKQAEKAAELSLAAEKATABEKEKASLLLEQKAAAEAEARQAAAAQAAEA 239
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNS 300
DQ |||||
Db 240 -----EKRAQQQSVLASANTNLTAQ-----VQAVSES 267
QY 301 DQESTTATEAOPSASSASTAVVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWAA 360
DQ |||||
Db 268 -----AAAPVRKVRPTVSTNASSYPICECTGWKSLAPWAGDYWGNGAQWAT 315
QY 361 SAAAGYRVGTPSPAGAVAVNDGGYGHVAVVTGQV-GQIQVQVANTAGNOSIGNYRGW 419
DQ |||||
Db 316 SAAAGYRTGTPPMVGAIAVNDGGYGHVAVVVEVQSSIRWESNYSGROYIADHRGW 375
QY 420 FNP-----GSVSYIY 429
DQ |||||
Db 376 FNPITTSSEGFYIY 390

RESULT 13
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
```

OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
FILING DATE: 10-Jul-2003
PRIORITY APPLICATION DATA:
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 3230:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...399
SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230

Query Match 50.2%; Score 1057.5; DB 5; Length 399;
Best Local Similarity 50.6%; Pred. No. 1.3e-56;
Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;
QY 1 MKKRILSAVLVSGVTLSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAVNTIQQ 60
DB 8 MKKILASLLSTVMSVQAVLTAAHTTDDKIAAQDNKISNLTAAQQAQAVNTIQQ 67
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 68 QVSAIQAEQSNLQAEQDRLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 127
QY 121 INAILNSKVSDAINRVSAIRVSVANEKMLQOQODKAAVEQKQOENQAATNTVAANO 180
DB 128 INTVNSKSIITSAIRKVAAMSEIVSANNKMLEQKQADKKAISEKQVANNDAINTVIANQ 187
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAQQKATLVAQKAAAEAAEAQAAQAAEA 240
DB 188 KLADDAQALTTQKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 246
QY 241 KAAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 300
DB 247 -----EKRAQQQSVLASANTLTAQ-----VQAVSES 274
QY 301 DQESTTATEAQSASSASTAVTANTSSANTYPAGCTGWKSLAPWVGNVWNGGOWAA 360
DB 275 -----AAAPVRKAVRPTYSNTASSYPIGCTGWKSLAPWVGNVWNGGOWAA 322
QY 361 SAAAGYRVGSPSAGAVVMDGGYGHVAYVTGQV-QQIQVQAEANYAGNOSIGNYRGW 419
DB 323 SAAAGYRVGSPSAGAVVMDGGYGHVAYVTGQV-QQIQVQAEANYAGNOSIGNYRGW 382
QY 420 FNP-----GSVSYIY 429
DB 383 FNPFTTSEGFTYIY 397

RESULT 14
US-10-282-122A-57658

Sequence 57658, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57658
LENGTH: 524
TYPE: PRT
ORGANISM: Enterococcus faecium
US-10-282-122A-57658
Query Match 28.4%; Score 598; DB 4; Length 524;
Best Local Similarity 31.2%; Pred. No. 1.9e-28;
Matches 163; Conservative 91; Mismatches 163; Indels 106; Gaps 14;
QY 1 MKKRILSAVLVSGVTLSATLTSVAKADDFDAQIASQDSKINNLTAAQQAQAVNTIQQ 60
DB 1 MKKILASLLSTVMSVQAVLTAAHTTDDKIAAQDNKISNLTAAQQAQAVNTIQQ 60
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 61 QVSEINTQAQDLAKQDTLRQESAQLVKDIADLQERIEKREDTIQKQAEQAQVNTSSNY 120
QY 121 INAILNSKVSDAINRVSAIRVSVANEKMLQOQODKAAVEQKQOENQAATNTVAANO 180
DB 121 IDAVLNADSLADIGRVQAVTTMTVKANNNDLMEQKQKQKQKQKQKQKQKQKQKQKQ 180
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAQQKATLVAQKAAAEAAEAQAAQAAEA 226
DB 181 ALESQKGDLLSKQADLNVLKTSLAABQATAEKADLNKQKAEAEQARIREQQRLAQ 240
QY 227 AARQAAAAQAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAA 281
DB 241 AQQAAQAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 300
QY 282 AAQTGDSSTEQSAAQAVNNNSDOESTTATEA-----QPSASSASTAVTANTSSANT 331

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:30:27 ; Search time 8.32046 Seconds
(without alignments)
736.166 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRGWFMFGSVIYPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_New.*

- 1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US05_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2092	99.4	431	7	US-11-052-554A-210 Sequence 210, App
2	1092.5	51.9	398	7	US-11-052-554A-252 Sequence 252, App
3	436	20.7	211	7	US-11-052-554A-352 Sequence 352, App
4	413.5	19.6	544	7	US-11-052-554A-358 Sequence 358, App
5	235	11.2	257	6	US-10-793-626-3244 Sequence 3244, App
6	227	10.8	971	7	US-11-052-554A-3 Sequence 3, Appl
7	226	10.7	394	7	US-11-052-554A-79 Sequence 79, Appl
8	225.5	10.7	1236	6	US-10-873-528-109 Sequence 109, App
9	215	10.2	149	6	US-10-793-626-1682 Sequence 1682, App
10	215	10.2	157	6	US-10-793-626-2870 Sequence 2870, App
11	215	10.2	270	6	US-10-485-517-413 Sequence 413, App
12	205	9.7	1562	7	US-11-052-554A-211 Sequence 211, App
13	198.5	9.4	2101	6	US-10-857-780-23 Sequence 23, Appl
14	196.5	9.3	1095	6	US-10-793-626-3154 Sequence 3154, App
15	192.5	9.1	1448	6	US-10-485-517-212 Sequence 212, App
16	189	9.0	1107	6	US-10-485-517-145 Sequence 145, App
17	187	8.9	758	6	US-10-485-517-144 Sequence 144, App
18	184	8.7	1586	6	US-10-821-234-901 Sequence 901, App
19	182	8.6	1410	6	US-10-878-556A-136 Sequence 136, App
20	180	8.6	1122	6	US-10-467-657-6112 Sequence 6112, App
21	179.5	8.5	330	6	US-10-485-517-415 Sequence 415, App
22	176.5	8.4	1126	6	US-10-485-517-248 Sequence 248, App
23	176	8.4	3712	7	US-11-019-711-48 Sequence 48, Appl
24	176	8.4	3712	7	US-11-019-711-51 Sequence 51, Appl
25	173	8.2	5024	6	US-10-793-626-2964 Sequence 2964, App

ALIGNMENTS

RESULT 1

US-11-052-554A-210
; Sequence 210, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 210
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-210

Query Match	99.4%	Score	2092	DB	7	Length	431
Best Local Similarity	99.3%	Pred. No.	7.6e-113				
Matches	428	Conservative	0	Mismatches	3	Indels	0
Gaps	0						
QY	1	MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIOG	60				
Db	1	MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIOG	60				
QY	61	QVSALQTQAEIQAENQRLAQSAITLGGQIQTLSSKIVARNESLKQARSAQKSNATSY	120				
Db	61	QVSALQTQAEIQAENQRLAQSAITLGGQIQTLSSKIVARNESLKQARSAQKSNATSY	120				
QY	121	INAINSKSVSDAIVNRVSAIREVVSANEKMLQQQODKAAVQKQENQAANTVAANOE	180				
Db	121	INAINSKSVSDAIVNRVSAIREVVSANEKMLQQQODKAAVQKQENQAANTVAANOE	180				
QY	181	TTAQTNTALNTQQAQLAAQLNLQALTTAODKATLVAQKAAAEAFQAAAAQAAAAA	240				
Db	181	TTAQTNTALNTQQAQLAAQLNLQALTTAODKATLVAQKAAAEAFQAAAAQAAAAA	240				
QY	241	KAAAEAKALQEAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDSFEQSAQAQVNNNS	300				
Db	241	KAAAEAKALQEAQAQAQAANNTQATDVSDDQAAAAADNTQAAQTGDSFEQSAQAQVNNNS	300				
QY	301	DQESTTATEAQPSSASSASTAVVVTANTSSANTYPAGCTWGVKSLAPFWGVNGVNGGQWAA	360				

Db 301 DQESTTATEAQSASASTAAVAANTSSANTYPAGCTGKSLAPWVGNWNGGQWAA 360
QY 361 SAAAGYRGVSTPSAGAVAVNDGGYGHVAYVTGQGGIQVQAEANYAGNQSIGNRGWF 420
Db 361 SAAAGYRGVSTPSAGAVAVNDGGYGHVAYVTGQGGIQVQAEANYAGNQSIGNRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match 51.9%; Score 1092.5; DB 7; Length 398;
Best Local Similarity 53.7%; Pred. No. 7e-56;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKGRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQOQAAQAVNTTIG 60
Db 1 MKGRILSAVLVSGVTILGAATT---VGAEDLSTKIAKQSIISNLTATTEQKAAQNOVSALQA 57

QY 61 QVSALOTQAELOAENORLEAOSATILGQOIOTLSSKIVARNESLKOQARSQKSAATSY 120
Db 58 QVSSIQSEQDKLTARNTLEALSKEPQEIKALTSGIVARNEKLNQAKSAYKNNETSGY 117

QY 121 INAINSKSVDAINRVSAIREVSNANEKWLQOQEQDKAAVQKQOENQAAINTVAANOE 180
Db 118 INALLNSKISDVNRLVAINRAVSANAKLLSQKADKVSLEKQAAQTAINTAANWA 177

QY 181 TTAQNTNALNTQOALEAQLNLQAEELTTAQQOKATLVAQKAAAEAAQAAQAAAE 240
Db 178 MAEENQTLTQANLEAATANLALQASATEDKANLVAQKAAAEAAEAALAEQAAKV 237

QY 241 KAAAEAKAQOQAAQAAQAAANNNTQATDASDQOQAAADNTQAAOTGDSSTEQSAAQVNS 300
Db 238 KAQEOA---AQQAASVEAAKSAITPAQA-----YDSNTYTPVQCTGAKSLAPWAGNNGGQWAA 280

QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGCTGKSLAPWVGNWNGGQWAA 360
Db 281 ----PAASARQTS-----YDSNTYTPVQCTGAKSLAPWAGNNGGQWAA 326

QY 361 SAAAGYRGVSTPSAGAVAVNDGGYGHVAYVTGQVQ-GGQIQVQAEANYAGNQSIGNRGW 419
Db 327 SAQAAGYRGVSTPMVGATAVNDGGYGHVAVVVEVQSASSIRVMESNSGRQYIADHRG 386

QY 420 NPGSVSYIYPN 431
Db 387 FNPTGVTFIYPH 398

RESULT 3

US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match 20.7%; Score 436; DB 7; Length 211;
Best Local Similarity 44.4%; Pred. No. 9.5e-19;
Matches 95; Conservative 28; Mismatches 57; Indels 34; Gaps 7;

QY 247 KALQOQAAQAAQAAANNNTQATDASDQOQAAADNTQAAQCTGDSSTEQSAAQ-----AV 297
Db 3 KOFLEKAVFTVAAA-----TAATVVLGNKMDAD-TYTLQEGDSFFSVAQRYHMDAYELASM 57

QY 298 NNSDOES-----TTATEAQPSPASASTAVVT-ANTSSANTYPAGCTGKVKSL 344
Db 58 NGKDTISLLPGQTLTVNGSAAPDNQAAAPDTTQATETNDANANTYPVGCTGKVKAV 117

QY 345 APWVGNWNGGQWAAASAAAAGYRGVSTPSAGAVAVNDGGYGHVAYVTG-OGGOIQVQ 403
Db 118 ATWAGDWNGGDMWASSASAQGYTVGNTPAVGSIMCMTDGGYGHVAYVTAVGEDGKVQL 177

QY 404 EANYAGNQSIGNYRGWFN-----PGSVSYIYPN 431
Db 178 ESNYKQDQWVDNVRGWFDPNNSGTGPGSVSYIYPN 211

RESULT 4
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match 19.6%; Score 413.5; DB 7; Length 544;
Best Local Similarity 25.4%; Pred. No. 5.4e-17;
Matches 143; Conservative 74; Mismatches 192; Indels 153; Gaps 17;

QY 4 RILSAVLVSGVTLSSTATTLSAVKADDFDAQTA-----SODSKINNLTAAQOQAAQAVNTIQ 59
Db 2 RKLKVALFASILGMLAVSSYTAADTEDNQVTISHYNEQAGTFDYNVQAANGKTIQSID 61


```

Qy 60 QGVSA LQTQQAELO-----AENORLEAQ-----ATL 86
   : : : :
Db 62 VAIWSEENGQDDLKWYHASNDGNSQLTWFHFAENHGSKVGSYIAHAYITTDGNRVGNL 121
   : : : :
Qy 87 GQOIQTLSSKIVARNESLKO--QARSACKNAATSYINAIINSKVSDAINRUSA--- 139
   : : : :
Db 122 GKRKLUS-----APQLSLUKQGLQLFSKLKPSAADQLFSVWSDENGDDLHWYTADADG 177
   : : : :
Qy 140 -----IREVVSANEKMLQOQEQD-----KAAVE-QKQEQENQAINTVAAN 178
   : : : :
Db 178 NTLAGYANHKGYGVTHVHTYLKONGKMPISAQDDIDIPKPKVKIQIDKINDTSDVWVNN 237
   : : : :
Qy 179 -----QETIAQNTNALNTQQAQLAAQLINQAE--- 206
   : : : :
Db 238 VPPYISSVAIPVWSEONGQDDLKWYQATKVADGIFKFTTVYLKTHRFELGNYQAHYGDQS 297
   : : : :
Qy 207 -----LTTAQDOKATI--VAQKAAAEAAARQAAAAQAAAAKAAEEA 246
   : : : :
Db 298 LSKKLDGLGETHFNVPISIINVEDPOVTTIDHYNINKGTFDVTVAETDNSKAIQISISAAVWS 357
   : : : :
Qy 247 KALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTG-----DSTEQSAAQ 295
   : : : :
Db 358 DANQANLYWYEAKLANGKAAITVDVQ-----KHGNGTGSYNVHVHVHYNDGTTSGHVL 411
   : : : :
Qy 296 AVNNSDQESTTATSAQPSASSASTAVTNTSSANTYPAGQCTWGVKSLAPVGVWYWGNG 355
   : : : :
Db 412 A-----NOQLNQIVHYQPSA-----VRITAYMNEKNTPVVGQCTWGVKSLAPWIPNLWNG 462
   : : : :
Qy 356 GOWASAAAAAGYRGVSTPSAGAVAVNNDGGYGHVAVYTVGVQ--GGQIQVQEAANYAGNQSTG 414
   : : : :
Db 463 GOWASTVAVKGFKLTGTPVKVGAIAICWSGGYGHVAVYTVHVSNNRIQVKEANYKQOYIS 522
   : : : :
Qy 415 NYRGWFNP-----GSVSYIYPN 431
   : : : :
Db 523 NFRGWFDPPTSYLGRLTIVYDP 544
   : : : :

```

RESULT 5

```

US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,358
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

```

Query Match 11.2%; Score 235; DB 6; Length 257;
Best Local Similarity 34.5%; Pred. No. 3.5e-07;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

261	Qy	NNNTQATDSDQOAAAADNTQAAQTGDSITEGSAQAQVNNDSQESITATEAQBPSSASTA	320
91	Db	NNYSNNYNNYQ---SNTQSQRTQTGGGLGASYSTSSNNVHVTTSA-PSSNGVSLU	146
321	Qy	VVTANTSSANTYPAGQCTWGY-KSLPWPVGNVYNGNGQWAAASAAAAGYRVGSTPSAGAVA	379
147	Db	--NARSAGNLVTSQCTTYTFDVRGGKIGSTGWGNANNWNAARASGYTVNNSPKAGAIL	204

Qy 380 VVNDGGYGHVAYVTGV-QGGQIQVQEANYAGNQSIGNYRGWFPNGSVSYIY 429
: ||||| || ||||| : ||||| : |||||
Db 205 OTSGAYGHVAYVEGVNNGSIRVSEMYGHGAGVVTSTISASQAAASYNY 255

RESULT 6

```

US-11-052-554A-3
; Sequence 3, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: prt
; ORGANISM: Escherichia coli 0157:H7
; US-11-052-554A-3

```

Query Match	10.8%	Score 227;	DB 7;	Length 971;
Best Local Similarity	23.8%	Fred. No. 4.5e-06;		
Matches	105;	Conservative	205;	Indels 56; Gaps 10;
Qy	19	ATTL\$AVKADDFDAQIASQDSKINLNTAQOQAAQAAQVNTIQGV\$ALOTQOAELOAENQR	78	
		::: :::: ::		
Db	121	AQNTAAKKSADAST\$AREAAATHATDAAD\$ARA\$AST\$AGAA\$ASSAG\$AST\$ATKA	180	
Qy	79	LEAQSATLGOOIQTLSGKIVARNESLKQOAR\$AQ\$KNAAT\$YINAININ\$KSV\$DAINRV\$	138	
Db	181	TEA-----\$K\$AAAE\$SK\$AAAT\$AGAA-\$AKT\$ETNAAV\$SOOG\$AA-----TS	221	

Qy		139	AIREVVSANEKMLQOOEQDAAVEKQOENQAATN-TVAANQETIAQN-----TNA	188
Dd	:	222	ASTATTTKASERASSARDASAKAEAKSSETSAASSASSAATAAGNSAKAAKTSETNA	281
Qy	:	189	LNTQQAQLE-----AAQLNLQELTTAQDQKATLVAQKAABEAARQAAAA--	234
Dd	:	282	KSESTAEEQSASAAAGSTAAALASAASASTSAGOASASATAAGKSAESAASSASTATTKA	341
Qy	:	235	-----QMAAEAKAEAEKALOEQAQAQAANNNTQTADSDQQAAAADNTQAAQTGDST	269
Dd	:	342	GEATPQAGAAASASAAKTSITNKASSETSBAESKTAASASSAASSASSASSASKDEAT	401
Qy	:	290	EQSAQAQAVNNDSQESTTATEAPQSASASASTAVVTANTSSANTYPAGQCTGWKSLAPWVG	349
Dd	:	402	RQASA-AKSATTASTKATEAGSATAAQSKSTAESAAATPAETAAKBAEDIAASALED	460
Qy	:	350	NYWGNGG--QMAASAAAAGYRVGSTPSGAVAVWNDDGGYGHVAYVTGVQGSGOIQVOEANY	407
Dd	:	461	ASTTKKGIVOLSSATNSTSESLAATPKRAKAAVELANGKYTAQDATTTAOKGIVOLSNATN	520
Qy	:	408	AGNQSIGNYRGWFNPGSVSYYIY	429
Dd	:	521	STSEMILA-----ATPKSVKAAAY	537

RESULT 7

```

RESULTS /
US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL ME
; TITLE OF INVENTION: PROTEINS OF THE

```

```
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226; DB 7; Length 394;
Best Local Similarity 27.3%; Pred. No. 1.8e-06;
Matches 108; Conservative 63; Mismatches 177; Indels 48; Gaps 15;

QY 5 ILSAVLVGVTLSATTLSAVKADDPDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQ 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 16 ILSAVL--HVILPALIWS-----FDENIEASAGGGSSIDA VWDVGAVVEQTKRMQ 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 GQVSALOTQQAELQAEORLEAQSATLGOQIOTLSSKIVARNESLKQ-----QARSAQKS 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 SQESS-----AKRSDQEKQEQAA-----BELREKQAAEQERLKOLEKERLAAEQEK 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 NAATSYINAINSKSVSDAINRVSAIREVVSANERKMLQOQEDKAAVEQKQOENQAINT 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 119 QAEAAKQAEKQKQAEAAKAAADAKAKAEADDKAAEAAKAAADAKKAAEAAKAA 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ---AELTTAQDQKATIVAQKAAAEAA--R 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 AEAQKKAEEAAALUKKGAEEAAEAEEAAEAARKKAAEAKAADKKA---AEKAAEAEEAA 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 QAAAAQAAAEKAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDST 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 KAAAEKAAADKKA--AKAAAEKAAAKAA--EADDIFGELSSGKNAPKTGGGA 287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 EQSAQAQVNSDQESTTATEAQ--PSASSASTAVTNTSSANTYPAGQCTWGVKSLAP-- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 KGNASPAAGSGNTKNGASGADINNYAGQIKSAIESKFYDASSYAGKCTTLRIK-LAPDG 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 347 WYGNWNGGQWA--ASAAAAGYRVGSTPSAGAVAVW 381
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 MLDIKPEGGDPALCOALAAAKIAKIPKPPSQAVY 382
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226; DB 7; Length 394;
Best Local Similarity 27.3%; Pred. No. 1.8e-06;
Matches 108; Conservative 63; Mismatches 177; Indels 48; Gaps 15;

QY 5 ILSAVLVGVTLSATTLSAVKADDPDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQ 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 16 ILSAVL--HVILPALIWS-----FDENIEASAGGGSSIDA VWDVGAVVEQTKRMQ 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 GQVSALOTQQAELQAEORLEAQSATLGOQIOTLSSKIVARNESLKQ-----QARSAQKS 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 SQESS-----AKRSDQEKQEQAA-----BELREKQAAEQERLKOLEKERLAAEQEK 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 NAATSYINAINSKSVSDAINRVSAIREVVSANERKMLQOQEDKAAVEQKQOENQAINT 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 119 QAEAAKQAEKQKQAEAAKAAADAKAKAEADDKAAEAAKAAADAKKAAEAAKAA 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ---AELTTAQDQKATIVAQKAAAEAA--R 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 AEAQKKAEEAAALUKKGAEEAAEAEEAAEAARKKAAEAKAADKKA---AEKAAEAEEAA 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 QAAAAQAAAEKAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDST 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 KAAAEKAAADKKA--AKAAAEKAAAKAA--EADDIFGELSSGKNAPKTGGGA 287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 EQSAQAQVNSDQESTTATEAQ--PSASSASTAVTNTSSANTYPAGQCTWGVKSLAP-- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 KGNASPAAGSGNTKNGASGADINNYAGQIKSAIESKFYDASSYAGKCTTLRIK-LAPDG 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 347 WYGNWNGGQWA--ASAAAAGYRVGSTPSAGAVAVW 381
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 MLDIKPEGGDPALCOALAAAKIAKIPKPPSQAVY 382
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match      10.2%; Score 215; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2.6e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 286 GDSTEQSAQAQVNSDQESTTATEAQPSASSASTAVTNTSSANTYPAGQCTWGV-KSL 344
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 APWVYNGWNGGQWAASAAAAGYRVGSTPSAGAVMNDGGYGHVAYVTGVGQ-GQIQVQ 403
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 68 GKGISTWGNANSWATAQAAGFTVNTPEGAIQSGEAGFGHVAFVESVNDGSIIVS 127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 404 EANYAG 409
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 EMNYDG 133
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Publication No. US20050272043A1
GENERAL INFORMATION:
APPLICANT: ROTH, RICHARD B.
APPLICANT: BRAUN, ANDREAS
APPLICANT: KAMMERER, STEFAN M.
APPLICANT: NELSON, MATTHEW ROBERTS
APPLICANT: RENELAND, RIKARD HENRY
APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
TITLE OF INVENTION: THEREOF
FILE REFERENCE: SEQ-4069-CP
CURRENT APPLICATION NUMBER: US/10/857,780
CURRENT FILING DATE: 2004-05-28
PRIOR APPLICATION NUMBER: 10/723,681
PRIOR FILING DATE: 2003-11-25
PRIOR APPLICATION NUMBER: 60/490,234
PRIOR FILING DATE: 2003-07-24
PRIOR APPLICATION NUMBER: 60/525,239
PRIOR FILING DATE: 2003-11-25
NUMBER OF SEQ ID NOS: 4962
SOFTWARE: PatentIn version 3.2
SEQ ID NO 23
LENGTH: 2101
TYPE: PRT
ORGANISM: Homo sapiens
US-10-857-780-23

Query Match 9.4%; Score 198.5; DB 6; Length 2101;
Best Local Similarity 22.6%; Pred. No. 0.00046;
Matches 96; Conservative 62; Mismatches 180; Indels 87; Gaps 14;
QY 2 KKEILSAVLVSGVTLSSATTLSA-----VKADDDFAQIASQDSKI-----NNLTAAQQA 51
DB 405 KGEVLGDLVLQLETLKQEAATLAANNLTQARVEMLETERGQOEAKLLAERGHFEKQOL 464
QY 52 QAQVNTIQGVSAQALQQAEL-----QAENQRLQAQATLGGQIQTLSSKIVARNE---SL 104
DB 465 SSLITDLOSSLNLSQAKLELQAQAHGARTAQVASITSELTINATIQDQQLAGL 524
QY 105 KQARSQAQSNATSYINAINSKSVDAINRVSAIREVVSANERKMLQOQEQODKAAVEOK 164
DB 525 KQAKKEQAQLAQT-----LQOQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEK 573
QY 165 Q-----QENQAANTVAANOETIAQNTALNTQAQLE---AAQLNL----- 203
DB 574 QBATQDHQAQATATAEEREASLRERDAALKQLEALEKEKAAKLETLQOQLQVANEARDS 633
QY 204 -QAEITTAQDQKATL---VAQKAAAEAEARQAAAAQAAAEAKAAAEAKALQEQAAQAA 259
DB 634 AQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQVAELQLRSEQQKATEKERV 693
QY 260 ANNTQ-----ATDASDQQAADNTQAAATQAAQTGDSSTEQSAA 294
DB 694 AQEKDQLOQLQALKESLKVTKGSLSEEEKRRADALEEQQRICSELK-AETRSLEQHKR 752
QY 295 QAVN-NSQESTTATEAQ-----PSASSASTAVV---TANTSSANTYPACQCTGWKSLAP 346
DB 753 EKKELEERAGRKGLLEARLLQIGEAHQAEIVRLRELAEMAAQHTASECEQQLKVEAA 812
QY 347 VVGNY 351
DB 813 WRDGY 817

RESULT 14
US-10-793-626-3154
Sequence 3154, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PUS480US
CURRENT APPLICATION NUMBER: US/10/793,626

CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3154
LENGTH: 1095
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-10-793-626-3154

Query Match 9.3%; Score 196.5; DB 6; Length 1095;
Best Local Similarity 22.0%; Pred. No. 0.00028;
Matches 96; Conservative 63; Mismatches 177; Indels 101; Gaps 12;
QY 13 GVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAAQVNTIQGOVSALQTOQAE 72
DB 340 GMTTDTANNYSKKREAE-ELQKAQIINNGDATEEQIITNTRVNQAINAKNDL 398
QY 73 QAENQRLQAQATLGGQIQTLSSKIVA--RNESLQQAARSQAQSNATSYINAINI 127
DB 399 RADKSOLENAYNQLIQNVDTNGKKPASIQOYQAARQAETQYNNAKSEAHQILENSPSV 458
QY 128 KSVSDAINRVSAIREVVSANERKMLQOQEQODKAAVEKQOENQA-----AINT 174
DB 459 NEVAQALQKVEAVQKVNDAIHMQLQNKNNLSALVTAKVQLQAVNDPFLTGTMTQDSINN 518
QY 175 VAA-----NOETIAQNTNA---LNTQQAQLEAAQLNLQ-----AELTTA 210
DB 519 YVAKNEAQAIRNAEAVINNGDATAKQISDEKSKVEQALAHNDKAKQOLTADTTTELQTA 578
QY 211 QDQ-----KATLVAQKAAAEAEARQAAAAQAAAEAKAAAEAKALQEQAAQAA 259
DB 579 VQOLNRRGDNTNKKPRSNAYNKAIQSLEQTITSAKDNANAVIQPIRTVQEVNNALQOV 638
QY 260 ANNTQATDASDQQAADN-----TQAQQTGDSSTEQS----- 292
DB 639 NQLNQLTEARINQLQPLSNNDALKAARLNLENKINQTVTQDGTQOOSIEAYQNAKRVQN 698
QY 293 ----AAQAVNNSD-----QESTTATEAQPSSASSASTAVVTANTSSANTYPA 334
DB 699 ESNTALALINNGDADEQIITETDRVNOQTNLITQAINGLTVNKEPLETEAKTALQNNDQ 758
QY 335 GQCTGWKSLAPWVGNV 351
DB 759 VPSTDGMTQQS--VANY 773

RESULT 15
US-10-485-517-212
Sequence 212, Application US/10485517
Publication No. US20050256299A1
GENERAL INFORMATION:
APPLICANT: University of Sheffield
APPLICANT: Biosynex Incorporated
APPLICANT: Foster, Simon
APPLICANT: Mond, James
TITLE OF INVENTION: Antigenic Polypeptides
FILE REFERENCE: P100629WO
CURRENT APPLICATION NUMBER: US/10/485,517
CURRENT FILING DATE: 2004-02-02
PRIOR APPLICATION NUMBER: GB 0118825.9
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: GB 0200349.9
PRIOR FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 424
SOFTWARE: PatentIn version 3.1
SEQ ID NO 212
LENGTH: 1448
TYPE: PRT

THIS PAGE BLANK (USPTO)